Modeling and estimating multivariate dependence structures with the Bernstein copula

Doro Rose

München 2015

Modeling and estimating multivariate dependence structures with the Bernstein copula

Doro Rose

Dissertation
an der Fakulät für Mathematik, Informatik und Statistik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Doro Rose
aus Würzburg

München, den 24. Juni 2015

Erstgutachter: Prof. Stefan Mittnik, PhD
Zweitgutachter: Prof. Sandra Paterlini, PhD
Tag der Disputation: 8. Oktober 2015

Abstract

In this thesis we study the properties of the Bernstein estimator for copulas and copula densities. Our aim is to clarify the relevance of this estimator for practical applications and to provide methodological enhancements. For practitioners questions regarding inference on copula models are the essential components in every analysis that is conducted within the copula framework. Naturally, since reliable inference can only be conducted on robust and valid information, considerable effort has been made to develop and test various methods of estimating copulas from empirical data. An influential contribution to this line of copula theory was published e.g. in 1995 by Genest et al. where the authors developed a semi-parametric estimation procedure for copulas that can be parameterized by a possibly vector valued dependence measure α.

Bernstein copula estimators were introduced in 2004 by Sancetta and Satchell. Based on Bernstein polynomials and their property to approximate any copula arbitrary well, Bernstein copulas constitute a very flexible and non-parametric tool to describe multivariate dependence structures. They are parameterized by the number of polynomials m that are applied to approximate the given copula, which in an estimation setting is given by the empirical copula. In a non-parametric framework the specification of an appropriate m is left to the analyst and to our knowledge, apart from cross-validation estimates, no explicit criterion that supports the latter with this task has been proposed yet.
We develop a data-driven estimator for m, which is optimal with respect to a criterion that is a functional of the mean squared error of the estimator. We compare our estimator in an extensive simulation study to the least square cross-validation method and show that our estimator is clearly superior to the latter in the bivariate case and that it generates tolerable estimates in 3 dimensional settings, even for small datasets. We develop the concept of local Bernstein copula estimation, that allows to specify the point estimator based on information that is local to the support of the respective dataset. Even though it was proved by Bouezmarni et al. in 2011 that the Bernstein copula estimator is asymptotically unbiased for unbounded copula densities, Bernstein estimates in the tail region of copula densities are usually poor in precision for small to medium sized samples, since realizations in these regions are sparse. Local Bernstein copula estimation allows us to enhance the precision on the total support and notably in the tails of the copula, furthermore this concept extends naturally to the non-parametrical analysis of dynamical copulas,

i.e. copulas that change in the course of time. An application of Bernstein copula estimators as model-selection tool is presented as well.

We apply these tools to financial datasets that are independent and identically distributed, as well as to datasets, that do not share this property.

Zusammenfassung

Bei der Analyse von Abhängigkeitsstrukturen zwischen Zufallsvariablen wird seit geraumer Zeit in verschiedenen Anwendungsbereichen, z.B. bei der Bewertung von Finanztiteln, auf Copulamodelle zurückgegriffen. Für den Anwender stellt sich neben der Auswahl des adäquaten Copulamodells auch die Frage nach der Inferenz, welche sich aufbauend auf den Ergebnissen der Schätzung ergibt.

Diese Doktorarbeit befasst sich mit den Eigenschaften des Bernsteinschätzers für Copulae und Copuladichten, welcher zu den nichtparametrischen Copulaschätzern zählt. Er basiert auf der Anwendung von Bernsteinpolynomen auf die empirische Copula. Bernsteinpolynome sind überwiegend aus der Approximationstheorie bekannt und wurden erstmals in 2004 von Sancetta undSatchell für die Anwendung auf Copulamodelle verwendet. Sie verhalten sich in allen wesentlichen Eigenschaften wie andere nichtparametrischen Schätzer, besitzen jedoch vorteilhafte Eigenschaften hinsichtlich ihrer asymptotischen Varianz. Das Ziel dieser Arbeit ist es die Eigenschaften dieses Schätzers für praktische Anwendungen zu analysieren und methodische Verbesserungen vorzuschlagen. In diesem Zusammenhang entwickle ich:

- einen semiparametrischen Schätzer für den Parameter, der die Bandbreite des Bernsteincopulaschätzers festlegt,
- das Konzept der lokalen Bernsteincopulaschätzung.

Derzeit wird der Bandbreitenparameter, der zur Spezifikation der Bernsteincopula notwendig ist in der Regel ad-hoc gesetzt, was sich erheblich auf die Güte der Schätzung auswirken kann. Durch die Entwicklung meines Schätzers biete ich eine robuste Alternative zu diesem derzeit verbreiteten Vorgehen. Die Robustheit des Schätzers weise ich u.a. in verschiedenen Simulationsstudien nach. Das Konzept der lokalen Bernsteincopulaschätzung bietet die Möglichkeit den Schätzer auf verschiedene Bereiche des untersuchten Datenraumes anzupassen, um so die Güte der Schätzung an diesen Stellen zu verbessern. Dies ist sinnvoll da die Schätzgüte des Bernsteincopulaschätzers in den Rändern des untersuchten Raumes nachlässt, letztere sind aber in vielen Anwendungen von besonderer Relevanz. Der lokale Bernsteincopulaschätzer ermöglicht zudem die Analyse von Abhängigkeitsstrukturen in Zeitreihendaten hinsichtlich deren zeitlicher Veränderung. Auch in diesem Kontext bieten sie den Vorteil gegenüber parametrischer Methoden, das sie ohne jegliche

Annahme bezüglich der tatsächlichen Copula auskommen. Eine Anwendung des lokalen Bernsteincopulaschätzers auf Finanzzeitreihendaten wird von mir gezeigt.

Acknowledgements

I would like to express my gratitude to the many people who have supported me directly and indirectly during the preparation of this thesis. Most of all I want to thank my supervisor Prof. Stefan Mittnik for encouraging me to embark on this venture, providing me with both, countless insightful advices along the way, as well as guidance in situations when I needed orientation. I would like to thank Prof. Sandra Paterlini for accepting to be co-examiner of this thesis.
Among the many that provided me with scientific, organizational or any other type of support that was necessary for me to complete this thesis I would like to mention the following people explicitly:

- My collegues at the Chair of Financial Econometrics at the Institute of Statistics of the LMU and especially to Fabian Spanhel, Andreas Fuest, Christian Groll and Serkan Yener.

- Martina Brunner for all the pleasant conversations, as well as patiently supporting me with every burocratic topic I adressed, which I can only guess must have been slightly annoying at some point. :-)

- My mother, sister and all my friends for their continuous support and motivation. I wouldn't have completed this thesis without their encouraging presence.

- Patricia for her patience, motivation and care ... I will always be thankful for that.

Contents

List of Figures

List of Tables

1. Introduction

Copulas can be thought of as multivariate probability distributions that are defined on the unit hypercube.

$$C\colon [0,1]^d \to [0,1]$$

This perspective differs slightly from the widespread interpretation that focuses on the fact, that a copula is the extracted dependency structure of a given probability distribution. Latter property is responsible for the great popularity of copula theory in applied probability theory as well as in real world applications, as it provides a very flexible way to construct joint probability distributions.[1][2]
Viewing a Copula as a probability distribution in its own right leads to the question of what can be inferred from such a probabilistic model. The information that is contained in the marginal distributions is stripped from hints to the quantitative extent of the random variables that generated the uniform margins and can solely be recovered by reverting the initial transformation. Thus the joint occurrences within a d-dimensional random vector $\mathbf{U} = (U_1, U_2, \ldots, U_d)$ with copula C are completely controlled by the copula function.

C assigns probabilities to the joint occurrences of the individual *ranks* of the random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)$, where

$$\mathbf{U} = (U_1, U_2, \ldots, U_d) = (F_1(X_1), F_2(X_2), \ldots, F_d(X_d)), \quad d \in \mathbb{N}.$$

F_i , $i \in \{1 \ldots d\}$ denote the respective cumulative distribution functions of $\mathbf{X}$. It is often helpful in a first step to analyze a copula solely with regards to the probability mass it assigns to the joint *ranks* and ignore the random variables that generated $\mathbf{U}$, as it visualizes the copula's equivalence to the rather abstract concept of "dependence structure". We are therefore able to state heuristically e.g. in the bivariate case

$$C\colon [0,1]^2 \to [0,1],$$

that a copula which assigns substantial probability mass along the diagonal in the $[0,1] \times [0,1]$ plane which passes through the pair $[(0.2, 0.2), (0.8, 08)]$ while allocating negligible probability mass on the counter-diagonal along the pair $[(0.8, 0.2), (0.2, 0.8)]$

[1] For applications in risk management see e.g. McNeil et al., 2010.
[2] For applications in finance see e.g. Cherubini et al., 2004.

represents a dependency structure that would most likely generate a positive correlation coefficient.[3] It should be emphasized, that the mechanism that steers this

(a) diagonal **(b)** counter-diagonal

Figure 1.1.: Orientation in the unit plane

interpretation relies heavily on the fact, that the margins of a copula are uniformly distributed on the unit interval.[4] To see this consider e.g. applications in Bayesian statistics where the uniform distribution is used as prior distribution in situations where ex ante there is no information available that allows the analyst to discriminate within the space of parameters of interest, thus equal probabilities are assigned to every possible outcome.[5] An equivalent interpretation in copula theory would state that the information with regards to the dependency structure that is contained in the marginal distributions is perfectly symmetrical and non-informative. It follows that by specifying any copula apart from the independence copula for a random vector $\mathbf{U}$ with pairwise independent entries, one explicitly induces a distinct dependence structure and removes the independence. This enables us to exercise a great amount of control when constructing relationships between random variables with regards to the specification of their joint occurrences. Considerations of causality within $\mathbf{U}$ are not yet addressed by copula theory.

For practitioners questions regarding inference on copula models become relevant after a copula has been extracted from a given dataset. The problem set that arises in the latter context is identical to the estimation problems that statistical theory has been dealing with traditionally. Given a sample of random multivariate data points the objective is to estimate the true copula that generated the sample. This can either be conducted in a fully parametrical procedure, where the necessary parameters of the margins and the copula can be estimated by maximum likelihood or by a semi-parametric approach. Latter is generally conducted in a two step procedure where the first step involves a transformation of the margins into uniformly

[3] Technical definitions of dependence concepts that can capture these properties precisely follow in chapter 2

[4] Other transformations of $\mathbf{X}$ are feasible and relevant in different contexts, see Embrechts, 2009.

[5] see e.g. Lehmann and Casella, 1998, Lehmann and Casella regard the usage of so called 'noninformative priors' as mimicking a state of ignorance with respect to the true state of nature.

distributed random variables $\mathbf{U} \sim (0, 1)$. If we possess the necessary information, we can use the true marginal distributions to perform this transformation, otherwise we have to use the empirical margins, which asymptotically converge to the true margins.[6] The second step comprises the actual estimation of the copula, which can be done either parametrically or non-parametrically. Kim et al. (2007) conduct an extensive simulation study to compare semi-parametric and parametric estimation methods for copulas. The authors find that semi-parametric methods are superior to fully parametric methods when the true margins are unknown, which is generally the case in practice. The Bernstein estimator for copulas represents a non-parametric approach to copula estimation problems that lends itself to semi-parametric estimation by following the steps outlined above.

This thesis is organized as follows. The copula function and its relation to dependence concepts is introduced in Chapter 2. In Chapter 3 we discuss central aspects and properties of the Bernstein copula, which is the main object of our analysis. We give details on possible estimation procedures and present two alternative representations, the multivariate baker representation and the Rook copula. This chapter concludes with an application related to model-selection. In Chapter 4 we lay the groundwork for the development of the Locally optimal Bernstein copula. We compare present approaches on bandwith selection and introduce a new criterion two estimate an optimal grid size for the Bernstein copula, penalized grid optimality. We conduct extensive simulation studies to assess the empirical behavior of the Bernstein copula estimator, parameterized with grid sizes specified according to their penalized grid optimality. Chapter 5 defines the concept of Locally optimal Bernstein copulas, a concept which allows the techniques developed in Chapter 4 to be adapted to various settings, such as estimation of the boarder regions of a copula or time series analysis.

[6] For a compact introduction to empirical process theory see e.g. Van der Vaart, 2000.

2. The Copula Function

As mentioned in the previous chapter a copula can be thought of as a d-dimensional distribution with uniformly distributed marginals $C : [0,1]^d \to [0,1]$ which satisfies the properties:

1. $C(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_d) = 0$,
2. $C(1, \ldots, 1, u, 1, \ldots, 1) = u$,
3. C is d-increasing, i.e for each hyper-rectangle:

$$B = \times_{i=1}^{d} [x_i, y_i] \subseteq [0,1]^d,$$

the C-Volume defined by:

$$\int_B \mathrm{d}C(u) = \sum_{\mathbf{z} \in \times_{i=1}^{d} \{x_i, y_i\}} (-1)^{N(\mathbf{z})} C(\mathbf{z}) \geq 0,$$

where the $N(\mathbf{z}) = \#\{k : z_k = x_k\}$ is non negative.

Property 1 establishes that zero probability mass is assigned to a copula,if one of its arguments is zero, whereas Property 2 establishes that a univariate margin may be recovered by integrating the copula over the other margins, which holds for multivariate distributions in general. The notion of d-increasing in Property 3 is merely the n-dimensional analog of a non-decreasing function of one variable, a necessary condition for a copula to be regarded as a multivariate probability distribution.[1] Two further results that are key in understanding the popularity of copula models are the quantile transformation and the probability transformation, which we present verbatim as in McNeil et al., 2010:

Proposition 1. *Let G be a distribution function and let $\overleftarrow{G}$ denote its generalized inverse, i.e. the function $\overleftarrow{G}(y) = inf\{x : F(x) \geq y\}$.*

1. ***Quantile transformation**. If $U \sim U(0,1)$ has a standard uniform distribution, then $P(\overleftarrow{G}(U) \leq x) = G(x)$.*

[1] see e.g. Nelsen, 2007, for a detailed introduction to the properties of a copula.

2. ***Probability transformation***. *If Y has df G, where G is a continuous univariate distribution function, then $G(Y) \sim U(0,1)$.*

For a proof we refer to the book of the authors or any introductory text in probability theory. The probability integral transform provides a simple way to transform a random variable into uniformly distributed random variables, whereas the quantile transformation enables to perfectly reverse the quantile transformation in the sense that the marginal distribution can be fully recovered. These transformations lie at the very heart of Sklar's theorem which is the central result within copula theory:[2]

Theorem 1. (Sklar 1959). *Let F be a joint distribution function with margins $F_1, \ldots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that, for all $x_1, \ldots, x_d$ in $\overline{\mathbb{R}} = [-\infty, \infty]$,*

$$F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right). \tag{2.1}$$

If the margins are continuous, then C is unique; otherwise C is uniquely determined on $Ran\, F_1 \times Ran\, F_2 \times \cdots \times Ran\, F_d$, where $Ran\, F_i = F_i(\overline{\mathbb{R}})$ denotes the range of F_i. Conversely, if C is a copula and $F_1, \ldots, F_d$ are univariate distribution functions, then the function F defined in 2.1 is a joint distribution function with margins $F_1, \ldots, F_d$.

Sklar's theorem states essentially, that it is possible to strip away the complete distributional information of the marginal distributions of a multivariate distribution and regain the information relevant to the multivariate dependence structure by means of the function C, that couples the uniformly distributed variables such that they are distributed like F. Thus a copula can be interpreted as dependence structure of a multivariate vector.

Another interesting relationship that should be focused is the relation of copula functions to scale invariant dependence measures.[3] If a relationship between a set of random variables was considered to be perfectly positively dependent, we would expect these variables to jointly increase and decrease in all possible states of the universe. We would expect the contrary to be the case if a set of random variables was considered to be perfectly negatively dependent. While the former is easily visualized in all dimensions the latter notion becomes conceptually difficult for dimensions $d > 2$. This heuristic notion of perfect positive dependence and perfect negative dependence is formalized by the Fréchet–Hoeffding bounds for copulas

[2] see McNeil et al., 2010, , for this translation.

[3] see Embrechts et al., 2002, , for a succinct analysis of the properties of Pearsons correlation coefficient and other dependence measures such as e.g. rank correlations.

given by

$$\max\{1-d+\sum_{i=1}^{d} u_i\} \leq C(u_1,\ldots,u_d) \leq \min\{u_1,\ldots,u_d\}, \tag{2.2}$$

where the lower bound constitutes a copula in the bivariate case which is commonly referred to as countermonotonicity copula

$$W(u_1,u_2) = \max(u_1+u_2-1,0).$$

In contrast the upper bound of Equation 2.2 constitutes a copula for all dimensions $d \geq 2$ and is commonly referred to as comonotonicity copula

$$M(u_1,\ldots,u_d) = \min\{u_1,\ldots,u_d\}.$$

In the bivariate case $W(\mathbf{u}$ and $M(\mathbf{u}$, where $\mathbf{u} = (u_1,u_2)$ are the distributions of the vectors $(U,1-U)$ and (U,U) respectively, as can be seen from

$$W(u_1,u_2) = P(U \leq u_1, 1-U \leq u_2), \tag{2.3}$$
$$M(u_1,u_2) = P(U \leq u_1, U \leq u_2), \tag{2.4}$$
$$\tag{2.5}$$

where $U \sim \mathcal{U}(0,1)$. $(U,1-U)$ represents a random vector that has all its probability mass on the counterdiagonal, i.e. between $(0,1)$ and $(1,0)$, whereas (U,U) represents a vector with all of its mass on the diagonal, i.e. between $(0,0)$ and $(1,1)$, see Figure 1.1. Thus the upper and lower Fréchet–Hoeffding bounds capture exactly the notion of perfect positive and perfect negative dependence as we described above. In between these two extremes lies the independence or product copula.

$$\prod(u_1,\ldots,u_d) = \prod_{i=1}^{d} u_i \tag{2.6}$$

which obviously represents independence. Figure 2.1 depicts these three copulas as well as their corresponding contour plots. To assess the height of each contour line it is easily seen from Equation 2 that the latter is identical to the value of u or v at the border of the graph.

In contrast to the notion of dependence that is captured by Pearson's correlation coefficient, $W(\mathbf{u})$ and $M(\mathbf{u})$ capture dependence via the complete distribution of the ranks of a pair of random variables $(F(X),F(Y))$. Embrechts et al., 2002 define a set of desired properties for a dependence measure $\delta(X,Y)$ that measures dependence

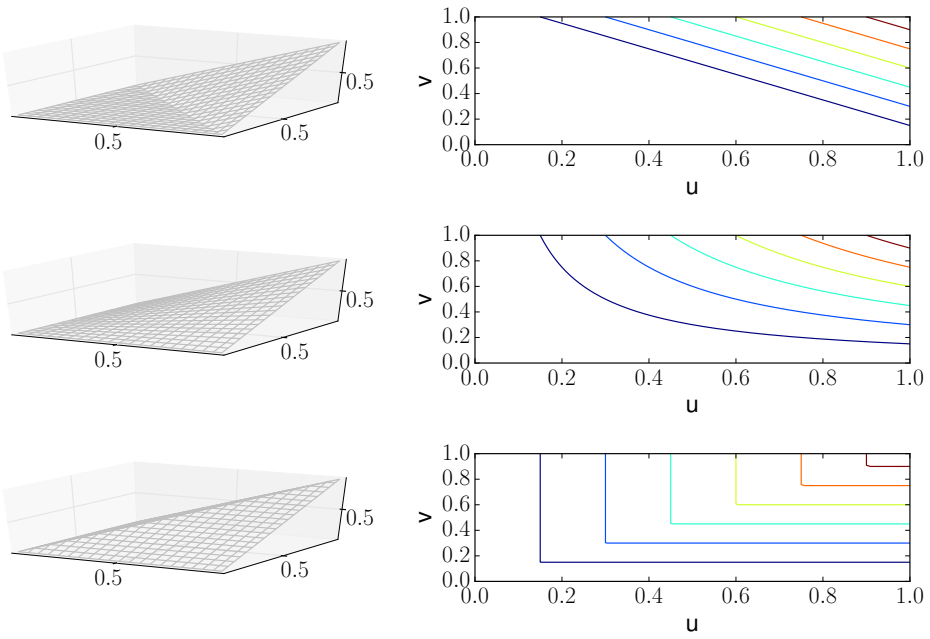

Figure 2.1.: First row: countermonotonicity copula, second row: product copula, third row: comonotonicity copula

in a single number as Pearson's correlation coefficient does, two of which being that

$$\delta(X,Y) = 1 \quad \Leftrightarrow X, Y \text{ comonotonic}, \tag{2.7}$$
$$\delta(X,Y) = -1 \Leftrightarrow X, Y \text{ countermonotonic}. \tag{2.8}$$

They further note that the natural requirement

$$\delta(X,Y) = 0 \Leftrightarrow X, Y \text{ independent} \tag{2.9}$$

contradicts the requirement that $\delta(X,Y)$ may possess the correct sign when (X,Y) show positive or negative dependence, i.e. for $T : \mathbb{R} \to \mathbb{R}$:

$$\delta(T(X),Y) = \begin{cases} \delta(X,Y), & \text{if } T \text{ increasing}, \\ -\delta(X,Y), & \text{if } T \text{ decreasing}. \end{cases} \tag{2.10}$$

They continue to show that rank correlations satisfy these and the other properties they have defined, which make them preferable to Pearson's correlation coefficient. The advantages of the former include invariance under monotone transformations, whereas Pearson's correlation coefficient is solely invariant under strictly increasing

linear transformations. We note that there is a copula representation for the two most widely known rank correlation coefficients Spearman's rho ρ_S and Kendall's tau ρ_τ:[4]

$$\rho_S = 12 \int_0^1 \int_0^1 [C(u,v) - uv]\, dudv, \tag{2.11}$$

$$\rho_\tau = 4 \int_0^1 \int_0^1 C(u,v) dC(u,v) - 1, \tag{2.12}$$

it is thus suitable to regard rank correlation coefficients as natural dependence parameters of copula functions. We will utilize this notion to develop an estimate for the bandwidth parameter of the Bernstein copula in Chapter 4. Another widely known measure that depends solely on the copula of a bivariate random vector are the coefficients of tail dependence given by

$$\lambda_u := \lambda_u(X_1, X_2) = \lim_{q \to 1^-} P(X_2 > \overleftarrow{F}_2(q) | X_1 > \overleftarrow{F}_1(q)), \tag{2.13}$$

$$\lambda_l := \lambda_l(X_1, X_2) = \lim_{q \to 0^+} P(X_2 \leq \overleftarrow{F}_q(q) | X_1 \leq \overleftarrow{F}_1(q)). \tag{2.14}$$

Their copula representation are

$$\lambda_l = \lim_{q \to 0^+} \frac{C(q,q)}{q}, \tag{2.15}$$

$$\lambda_u = \lim_{q \to 1^-} \frac{\hat{C}(1-q, 1-q)}{1-q} = \lim_{q \to 0^+} \frac{\hat{C}(q,q)}{q}, \tag{2.16}$$

$$\tag{2.17}$$

where $\hat{C}$ is the survival copula of C.[5] The variables X_1, X_2 are said to show upper or lower tail dependency if the respective coefficient is in $]0,1]$, if $\lambda_u = 0$ or $\lambda_l = 0$ they are considered asymptotically independent in the respective tail.

[4] Copula representations for two lesser known rank correlation coefficients Gini's gamma and Blomqvist's beta are given by:

- $\gamma_{Y_1,Y_2} = 4 \int_0^1 [C(u_1, 1-u_1) + C(u_1, u_1)) du_1 - 1]$
- $\beta_{Y_1,Y_2} = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1.$

[5] see McNeil et al., 2010, for further details on the tail dependance coefficients.

3. The Bernstein Copula

The ability of Bernstein polynomials to approximate any function $f(x)$ arbitrarily well has been well known in approximation theory.[1] They were developed by Bernstein, 1912 to provide a constructive proof of the Weierstrass Approximation Theorem. Their application to copula approximation has been studied by Li et al., 1997, Li et al., 1998, Kulpa, 1999 who focus on convergence properties of Bernstein and so called *checkerboard* copulas, as well as Durrleman et al., 2000b and Durrleman et al., 2000a who established the convergence of Spearman's rho and Kendall's tau to their approximands for both *checkerboard* and Bernstein copulas. These results were established in bivariate settings. Applications to frequentist density estimation have been introduced by Vitale, 1975 for univariate densities and Tenbusch, 1994 for bivariate densities, while their applicability within a Bayesian framework has been investigated by Petrone, 1999a, Petrone, 1999b and Petrone and Wasserman, 2002. Bernstein polynomials have been discussed within regression analysis by Tenbusch, 1997 and Brown and Chen, 1999. Their properties with regard to the estimation of copula functions were analyzed by Sancetta and Satchell, 2001 and Sancetta and Satchell, 2004, as well as Bouezmarni et al., 2008, Bouezmarni et al., 2010 and Bouezmarni et al., 2013. Janssen et al., 2012 provide the almost sure consistency and asymptotic normality for the empirical Bernstein copula density estimator and an analysis with respect to vine copula constructions is presented in Weiss and Scheffer, 2012. For a recent publication on Bernstein copulas see Dou et al., 2014 or Cottin and Pfeifer, 2014, both of which introduce alternative representations of Bernstein copulas, which we will discuss in the Sections 3.4.1 and 3.4.2.

3.1. Definition

Let $f(x)$ be a function defined on the interval $[0, 1]$, then

$$B_n(x) = B_n^f(x) = \sum_{\nu=0}^{n} f\left(\frac{\nu}{n}\right) \binom{n}{\nu} x^\nu \left(1-x\right)^{n-\nu}, \quad \nu \in \{0, \ldots, n\}.$$

[1] see Lorentz, 2012, for an exposition on the main facts of Bernstein polynomials.

is called the Bernstein polynomial of order n of the function $f(x)$.[2] The n + 1 expressions

$$b_{\nu,n}(x) = \binom{n}{\nu} x^{\nu} (1-x)^{n-\nu}, \quad \nu \in \{0, \ldots, n\}. \tag{3.1}$$

are also referred to as Bernstein basis polynomials. In general any linear combination of Bernstein basis polynomials

$$B(x) = \sum_{\nu=0}^{n} \beta_{\nu} b_{\nu,n}(x),$$

is called a Bernstein polynomial or polynomial in Bernstein form of degree n. The coefficients β_{ν} are called Bernstein coefficients or Bèzier coefficients. Let,

$$P_{m_i,v_i}(u) = \binom{m_i}{v_i} u^{v_i} (1-u)^{m_i - v_i}, \quad v_i \in \{0, \ldots, m_i\},$$

Let $\alpha\left(\frac{v_1}{m_1} \ldots, \frac{v_d}{m_d}\right)$ be a real valued constant indexed by $(v_1, ..., v_d)$, $\quad 0 \leq v_j \leq m_j$. If $C_B : [0,1]^d \to [0,1]$, where

$$C_B(u_1, \ldots, u_d) = \sum_{v_1} \ldots \sum_{v_d} \alpha\left(\frac{v_1}{m_1}, \ldots, \frac{v_d}{m_d}\right) \prod_{i=1}^{d} P_{m_i,v_i}(u_i), \tag{3.2}$$

satisfies the properties of the copula function, then C_B is a Bernstein copula for any $m_j \geq 1$. Each margin u_j of the Bernstein copula is separated into $m_j, \quad j = 1, \ldots, d$ sections, giving rise to a grid-type structure in the d-dimensional hypercube. If $m = m_j, \quad \forall j = 1, \ldots, d$ we will refer to m as the *grid*-size of the Bernstein copula. C_B is evaluated at each vertex of this grid by its coefficient, the constant $\alpha\left(\frac{v_1}{m_1}, \ldots, \frac{v_d}{m_d}\right)$, which is then smoothed by the product of Bernstein polynomials $\prod_{i=1}^{d} P_{m_i,v_i}(u_i)$. The latter has the analogous interpretation as the kernel function in kernel density estimation, i.e. it can be thought of as weight function that adjusts the constant α by applying information from the whole dataset, with the difference being that the weight function is adjusted according to the position of its arguments. The constant $\alpha\left(\frac{v_1}{m_1}, \ldots, \frac{v_d}{m_d}\right)$ represents the value of the estimated copula, evaluated at the a specific vertex of the grid. In the case of estimating an unknown copula by its empirical copula α is recovered from the dataset by counting the number of observations in the respective hypersquare of the grid, the boundary of which is specified by the arguments of α. Figure 3.1 depicts the smoothing effect

[2] see Lorentz, 2012.

of 2-dimensional Bernstein polynomials. Each color represents one polynomial, the graphs are generated by fixing one value of the binomial coefficient and evaluating the product $\prod_{i=1}^{2} P_{m_i,v_i}(u_i)$, for all values of u, v on $[0,1]^2$. We can see that at each point in the plane the weight that is assigned to the respective coefficient is a result of the overlap of multiple polynomials. We can further observe that, as mentioned above, the structure of the weight depends on the position of the point.

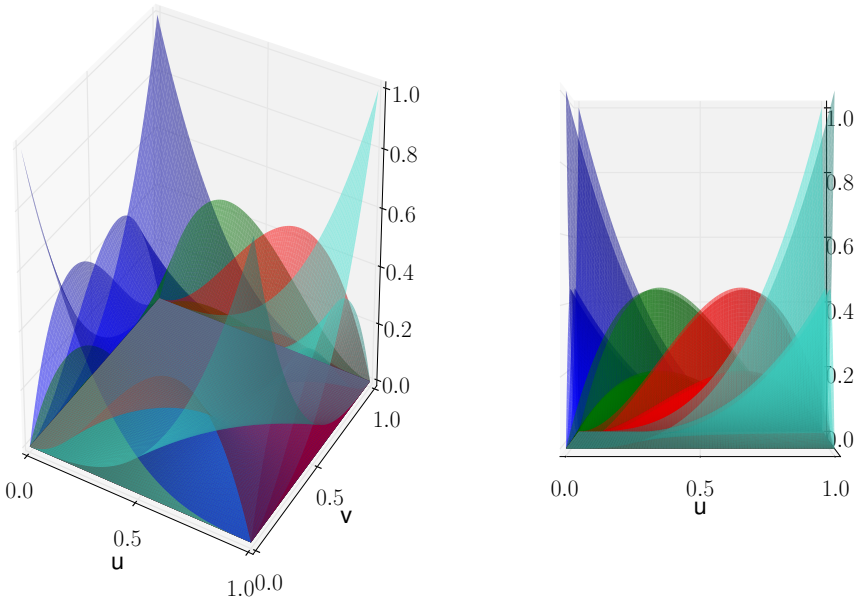

Figure 3.1.: Smoothing effect of 2-dimensional Bernstein polynomials: m=3

The Bernstein copula can be thought of as a generalization of the class of polynomial copulas, i.e. the class of copulas that can be expressed as polynomials in at least one of the margins of $\mathbf{u}$. Their general form in the bivariate case is

$$C(u,v) = a_n(v)u^n + a_{n-1}(v)u^{n-1} + \cdots + a_1(v)u + a_0(v),$$

with appropriate functions $a_0, \ldots, a_n$. Members of this class are e.g. the Product copula 2.6 mentioned in Chapter 2, which has linear sections in all variables or the bivariate Farlie-Gumbel-Morgenstern Copula:

$$C_\theta = uv + \theta uv(1-u)(1-v), \tag{3.3}$$

which is quadratic in both sections.[3]

[3] see Nelsen, 2007, for a discussion of polynomial copulas.

The Bernstein copula as we introduced it in Equation (3.2) relates directly to approximation theory as described in Durrleman et al., 2000a and Li et al., 1997 . In this context the problem that arises is the construction of a copula $\mathbf{C}$ from its discretization, i.e. from the information provided on the grid $\mathcal{L} = \left\{\left(\frac{i}{m}, \frac{j}{m}\right) : 0 \leq i, j \leq 1\right\}$, on which the value of $\mathbf{C}$ is known at every vertex. The central point in this perspective is that for the approximation to be valid, it has to converge to the approximand, the underlying copula, as m increases. At this stage there are two perspectives that arise naturally, the first being the statistical perspective in which the discretization can be viewed as an empirical observation of the unknown copula, the second being the pure approximation perspective, which results in the aforementioned construction problem of a valid copula, since there are infinitely many copulas that can be related to a specific discretization. As a solution to the latter the authors construct a bijection between the set $\mathcal{D}_n$ of *doubly-stochastic* n*x*n *matrices* and the set $\mathcal{C}_n$ of *discretized* copulas at each point $\left(\frac{i}{n}, \frac{j}{n}\right)$ of the grid $\mathcal{L}$:

$$\omega : \quad \mathcal{D}_n \to \mathcal{M}_{n+1} \tag{3.4}$$

$$A \longmapsto B = \omega(A) \tag{3.5}$$

with

$$b_{0,i} = b_{i,0} = 0 \quad 0 \leq i \leq n \tag{3.6}$$

and

$$b_{i,j} = \frac{1}{n} \sum_{p=1}^{i} \sum_{q=1}^{j} a_{p,q} \quad 1 \leq i, j \leq n \tag{3.7}$$

Thus a *doubly-stochastic* matrix can be viewed as a discrete representation of a bivariate copula density. Figure 3.2 depicts a 4x4-*doubly-stochastic* matrix $\mathcal{D}_4(\mathbf{C})$ and a discretized bivariate Gauss copula $\mathcal{C}_4(\mathbf{A})$. The map from the *doubly-stochastic* matrix to the discretized copula was constructed according to Equation (3.7).

If one sets $\alpha = \mathcal{C}\left(\frac{v_1}{m_1}, \ldots, \frac{v_d}{m_d}\right)$ in Equation (3.2) the unknown copula $\mathcal{C}$ is evaluated at discrete sections defined by its arguments, thus in an empirical context α can be thought of as discrete approximation of the unknown copula. Pfeifer et al., 2009 generalize this construction by defining so called *grid-type* copulas, which subsume all copulas, that are defined on a grid.

For the majority of this thesis we will be focusing on the Bernstein copula density, which is given by:

$$c_B(u_1, ..., u_k) = \prod_{i=1}^{d} m_i \sum_{v_1} ... \sum_{v_d} \alpha\left(\frac{v_1}{m_1} \ldots, \frac{v_d}{m_d}\right) \prod_{i=1}^{d} P_{m_i, v_i}(u_i), \tag{3.8}$$

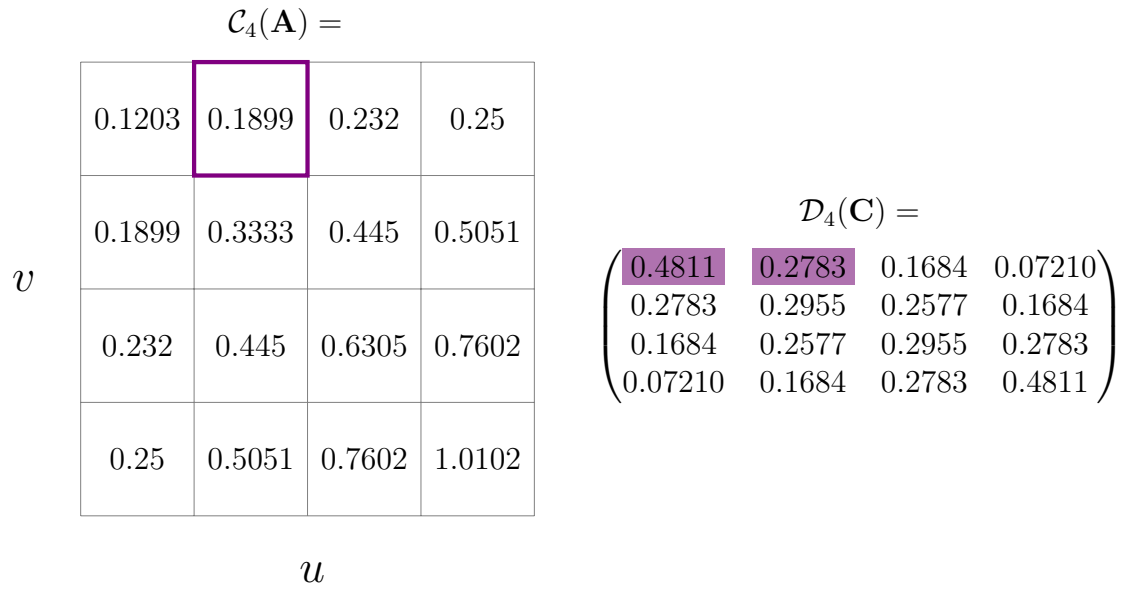

Figure 3.2.: Correspondence between doubly stochastic matrix and a bivariate Gauss copula with $\rho = 0.5$ and $\sigma = 1$. The colored region corresponds to the sum of the matrix elements a_{11} and a_{12}, divided by 4: $\frac{a_{11}+a_{12}}{4}$

3.2. Properties

In the following section we collect some basic properties of the Bernstein copula and illustrate how random samples that follow a specific Bernstein copula can be generated. The asymptotic behavior of the Bernstein copula is driven by the unique properties, that Bernstein polynomials possess with regard to the approximation of arbitrary functions.

3.2.1. Large sample behavior

The main developments regarding the asymptotic properties of the Bernstein copula density estimator were made by Sancetta and Satchell, 2004 and Bouezmarni et al., 2010. Former construct the estimator for the copula function and find an upper bound for the asymptotic bias and variance and establish the asymptotic normality of the Bernstein copula density estimator in an i.i.d. setting. The latter extend this analysis to α-mixing dependent data and provide the exact asymptotic bias and variance, as well as establishing uniform strong consistency and asymptotic normality. In Bouezmarni et al., 2013 the authors prove that the Bernstein copula density estimator converges to infinity at the boarders, as well as providing further convergence and consistency results. These results extend the results established in

Bouezmarni et al., 2010 and Sancetta and Satchell, 2004 to the estimation of copula densities that are unbounded, as latter results were dependent on a boundedness assumption of the underlying copula. We present the main results of this line of research in the present section.
For a sequence to be α-mixing the following has to hold:

$$\alpha(h) = \sup_{A\in\mathcal{F}_1^t(X), A\in\mathcal{F}_{t+h}^{\infty}(X)} |P(A\cap B) - P(A)P(B)|, \tag{3.9}$$

where $\mathcal{F}_1^t(X)$ and $\mathcal{F}_{t+h}^{\infty}(X)$ are the σ-field of events generated by $\{X_l, l \leq t\}$ and $\{X_l, l \geq t+h\}$, respectively. Bouezmarni et al., 2010 further require

$$\alpha(h) \leq \rho^h, \quad h \geq 1, \tag{3.10}$$

for some constant $0 < \rho < 1$.

For a sequence of α-mixing dependent random vectors the asymptotic bias of the Bernstein copula density estimator $\hat{c}$ is then given by

$$\mathrm{E}(\hat{c}_B(\mathbf{u})) = c(\mathbf{u}) + \frac{\gamma^*(\mathbf{u})}{m}\mathcal{O}\left(m^{-1}\right) \tag{3.11}$$

where

$$\gamma^* = \frac{1}{2}\sum_{j=1}^{d}\left\{\frac{dc(\mathbf{u})}{du_j}(1-2u_j) + \frac{d^2c(\mathbf{u})}{du_j^2}u_j(1-u_j)\right\}. \tag{3.12}$$

The asymptotic variance for $\mathbf{u} \in]0,1[^d$ is given by:

$$\mathrm{Var}(\hat{c}_B(\mathbf{u})) = \frac{m^{\frac{d}{2}}V(\mathbf{u})}{n} + \mathcal{O}\left(\frac{m^{\frac{d}{2}}}{n}\right) \tag{3.13}$$

where

$$V(\mathbf{u}) = 4\pi^{-\frac{d}{2}}\frac{c(\mathbf{u})}{\prod_{j=1}^{d}\left(u_j(1-u_j)\right)^{\frac{1}{2}}}, \tag{3.14}$$

and at the boundary $\mathbf{u} = 0, \mathbf{u} = 1$ by:

$$\mathrm{Var}(\hat{c}_B) = \frac{m^d}{n}c(u) + O(\frac{m^{d-1}}{n}) \tag{3.15}$$

From the denominator in Equation (3.14) we see that the variance of $\hat{c}_B$ increases near the boundary, nevertheless, as mentioned above, the estimator is in a sense precise as relative convergence at the corners as well as convergence to infinity, when the density is unbounded, was proven. In general we thus conclude, that

1. $\hat{c}_B$ is asymptotically unbiased.
   $Bias(\hat{c}_B) = O(m^{-1})$, $m \to \infty$.

2. The variance of $\hat{c}_B$ vanishes asymptotically on the entire support.
   At the boundary it vanishes if $\frac{m^d}{n} \to 0$.
   In the interior support it vanishes if $m^{\frac{d}{2}} \to 0$.

3. The optimal grid size in a mean square error sense increases with rate $O(n^{2/m+4})$.

4. From 1. and 2. it follows that $\hat{c}_B$ is a consistent estimator for the estimated true copula density, i.e. $\hat{c}_B(\mathbf{u}) \to c(\mathbf{u})$.

Bernstein polynomials have slower rate of convergence compared to other polynomial expressions, but they perform best within the class of all operators with similar shape preserving properties.[4][5] In principle the slow rate of convergence can be compensated by increasing the grid size m, but from Fact 2. we can immediately see, that while this would decrease the bias of the estimator it has the drawback of simultaneously increasing its variance. A MSE-optimal grid size would take account of this trade off, see Chapter 4. For the following comparisons between the Bernstein estimator and the kernel and histogram estimators we follow Sancetta and Satchell, 2004 and set the usual bandwidth parameter $h \equiv m^{-1}$. Compared to the kernel estimator, which has a bias of order $\mathcal{O}(m^{-2})$, both the Bernstein and histogram estimators have a higher bias of order $\mathcal{O}(m^{-1})$, but the variance of the Bernstein estimator is lower since the former and the histogram estimator have variance of order $\mathcal{O}(m^d)$, compared to $\mathcal{O}(m^{\frac{d}{2}})$ for the Bernstein estimator inside the hypercube. Combining this information we could focus on the rate of consistency of the kernel, histogram and the Bernstein estimators and observe that for the former $\hat{c}_B(\mathbf{u}) \to c(\mathbf{u})$ holds at rate $\frac{m^d}{n} \to 0$, while the same holds for the latter inside the hypercube at $\frac{m^{\frac{d}{2}}}{n} \to 0$. The optimal order of smoothing with regard to the MSE-criterion for the Bernstein estimator is $m = \mathcal{O}(n^{\frac{2}{d+2}})$, $m = \mathcal{O}(n^{\frac{1}{d+2}})$ for the histogram estimator and $m = \mathcal{O}(n^{\frac{1}{d+4}})$ for the kernel estimator. Thus the Bernstein estimator requires less smoothing compared to the other two non-parametric estimators.

As mentioned in the previous chapter rank-correlations and tail dependency can be regarded as natural dependence measure for copula functions, consequently their exist various representations for Spearman's rho and Kendall's tau for the Bernstein copula. Durrleman et al., 2000a provide the following elegant representations

[4] see Sancetta and Satchell, 2004.
[5] see Berens and DeVore, 1980.

Theorem 2. (Durrleman et al., 2000a) *Let* $\Theta \in M_{n+1}$ *where*

$$\theta_{i,j} = \frac{(i-j)\binom{n}{i}\binom{n}{j}}{(2n-i-j)\binom{2n-1}{i+j-1}}$$

for all $0 \leq i, j \leq n$ *(with the convention 0/0 = 1), then the Kendall's tau of the Bernstein copula is*

$$\tau = 1 - \operatorname{tr}\left(\Theta D \Theta D^{\top}\right)$$

where $D \in \mathcal{C}_n$ denotes an element of the set $\mathcal{C}_n$ of discretized copulas at each point $\left(\frac{v_1}{m_1}, \frac{v_2}{m_2}\right)$ of the grid $\mathcal{L}$ generated by the grid size m, i.e. D should be thought of as the copula counterpart of the *doubly-stochastic* matrices introduced in the previous chapter.

Theorem 3. (Durrleman et al., 2000a) *Let* $\Gamma \in \mathcal{M}_{n+1}$ *where*

$$\gamma_{i,j} = \frac{1}{(n+1)^2}$$

for all $0 \leq i, j \leq n$ *(with the convention 0/0 = 1), then the Spearman's rho of the Bernstein copula is*

$$\rho = 12 \operatorname{tr}(\Gamma D) - 3$$

The Bernstein copula does not preserve tail dependency.[6][7] Nevertheless we regard this as a minor drawback since the Bernstein copula is perfectly capable of capturing dependency that exists in a tail, it is merely the limiting operation that fails to be replicated, if the underlying copula exhibits tail dependency.

3.2.2. Random number generation

To generate a random sample from a Bernstein copula two methods have been proposed in the literature, the first being the application of rejection sampling, whereas the second is based on the fact, that a Bernstein copula can be expressed as a mixture of Beta distributions.
Bernstein copula densities are polynomials, which are evaluated on the unit hypercube $[0,1]^d$, thus they are bounded by a constant $M > 0$. The necessary steps to generate random variables with the multivariate acceptance-rejection method are provided by Pfeifer et al., 2009 as followed:

1. Generate $d+1$ independent uniformly distributed random numbers $u_1, \ldots, u_{d+1}$.

[6] see Sancetta and Satchell, 2004, for a heuristic argument to this property.
[7] see Durrleman et al., 2000a, for a proof with regard to λ_u.

2. If $c(u_1, \ldots, u_d) > M u_{d+1}$ go to step 3, otherwise go to step 1.

3. Use $(u_1, \ldots, u_d)$ as a sample from the Bernstein copula.

Consequently M repetitions of this algorithm are necessary on average to generate 1 sample from the Bernstein copula. This method has the drawback of not being feasible for high dimensional problems, since the rejections increase due to the curse of dimensionality, i.e. the ratio between the volume of the acceptance region to the rejection region tends towards zero. Furthermore, upper bounds are not generally easily obtained in high dimensional settings. If there is a preference to generate random samples via Monte Carlo methods one would consequently have to switch to different algorithms such as the Metropolis-Hastings algorithm or Gibbs sampling.[8]

The second method, which was proposed by Diers et al., 2012 and Cottin and Pfeifer, 2014, relies on the representation of Bernstein copula densities as mixtures of Beta distributions. Let $\mathbf{Y} = (Y_{v_1,m_1}, \ldots, Y_{v_d,m_d})$ denote a vector of Beta distributions with parameters $(v_j + 1, m_j - v_j)$, $j = 1, \ldots, d$. The respective densities are then given by

$$\begin{aligned} f_{Y_{v_1,m_1}}(z) &= m_j \binom{m_j - 1}{v_j} z^{v_j} (1-z)^{m_j - 1 - v_j} \\ &= \frac{1}{B(v_j + 1, m_j - v_j)} z^{v_j} (1-z)^{m_j - 1 - v_j}, \end{aligned} \tag{3.16}$$

where $z \in [0, 1]$.
The algorithm for generating random samples via the Beta mixture representation is then given by:

1. Generate a random sample $(V_1, \ldots, V_d) \in [0, \ldots, m-1]^d$ such that $P[(V_1, \ldots, V_d) = (v_1, \ldots, v_d)] = \alpha(\frac{v_1}{m}, \ldots, \frac{v_d}{m})$.

2. Sample $(U_i, \ldots, U_d)$ with independent $U_i \sim Beta(V_i, m_i - V_i)$ for $i = 1, \ldots, d$.

3.3. Estimation

Compared to the vast literature on parametrical copula estimation, the literature on non-parametrical copula estimation is rather negligible. Contributions in this regard have dealt with kernel density estimation as proposed by Gijbels and Mielniczuk, 1990, Fermanian, n.d. who extended the analysis to time-dependent data, Fermanian et al., 2004 and Chen and Huang, 2007. Omelka et al., 2009 improved the kernel estimator by correcting for boundary bias. Wavelet based copula density estimation

[8] see Devroye, 1986, for a comprehensive overview of sampling methods.

was studied in Genest et al., 2009 and Autin et al., 2010, whereas Morettin et al., 2010 analyze copula estimation. Qu et al., 2009 and Qu and Yin, 2012 estimate the copula density by smoothing a piecewise constant function placed on a grid and maximizing a penalized log likelihood, where the penalty is incorporated to impose smoothness on the estimate. Finally, non-parametric estimation based on the application of B-splines was suggested by Shen et al., 2008, Kauermann et al., 2013 and Cormier et al., 2014. For an overview of recent approaches we refer to Embrechts, 2009 and Jaworski et al., 2010.

3.3.1. Non-parametric estimation

A fully non-parametric approach of estimating the Bernstein copula density $\hat{c}_B$ consists of two steps. In the first step the probability transformation is applied to each element of the random vector $(X_1, \ldots, X_d)$ by means of their respective empirical distributions $\left(\hat{F}_1, \ldots, \hat{F}_d\right)$ and in the second step the elements of the vector $\hat{F}_1(X_1), \ldots, \hat{F}_d(X_d)$ the Bernstein copula density is estimated via the empirical copula density, which is identical to a multivariate histogram estimator.[9][10] The quality of the approximation of the true marginal distributions hinges obviously at the size of the sample under consideration, as we know from empirical process theory.[11] Thus these approximations get better as the sample size increases. In addition to this it is not guaranteed that the marginal distribution generated by the grid which is applied to the joint relative ranks satisfy the predicate of being uniformly distributed. To see this we look at Figure 3.3, which depicts 100 2-dimensional samples from a Clayton copula with $\theta = 0.5$ upon which we apply a 10x10 grid. The corresponding joint relative ranks are depicted in the upper table of Table 3.1. We can observe that the margins are not perfectly uniformly distributed, as can be seen from the row and column sums, even though the samples were drawn from the Clayton copula, i.e. the margins are uniformly distributed by definition. It is the rows and columns we have to asses, since the Bernstein copula evaluates $P(m_1U_1 \leq v_1, \ldots, m_dU_d \leq v_d)$,

[9] see Scott, 2012.

[10] The empirical copula density is sometimes also referred to as *joint relative ranks*, see e.g. Pfeifer et al., 2009

[11] With regard to copula estimation Deheuvels, 1979, were the first to consider the empirical copula. They furthermore proved the weak convergence of the empirical copula process, when the marginals are independent.

Table 3.1.: Joint relative ranks, original data vs optimized data. The first table shows the original data, which was sampled from a Clayton copula with $\theta = 0.5$. The entries $a_{i,j}$ are not perfectly uniformly distributed, as can be seen from the last row and column, which depict the respective sums. In contrast, the entries $x_{i,j}$ of the optimized dataset are much closer to being uniformly distributed.

		u										
	$a_{ij} \leq$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	$\sum$
v	1.0	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.02	0.00	0.04
	0.9	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.04	0.07
	0.8	0.00	0.00	0.00	0.03	0.03	0.02	0.00	0.02	0.02	0.00	0.12
	0.7	0.00	0.02	0.00	0.04	0.00	0.02	0.03	0.01	0.01	0.02	0.15
	0.6	0.00	0.01	0.00	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.12
	0.5	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.02	0.02	0.01	0.10
	0.4	0.00	0.01	0.04	0.01	0.00	0.01	0.01	0.02	0.01	0.00	0.11
	0.3	0.00	0.00	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.10
	0.2	0.03	0.01	0.01	0.00	0.01	0.02	0.00	0.01	0.00	0.00	0.09
	0.1	0.07	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.10
	$\sum$	0.10	0.07	0.08	0.13	0.10	0.12	0.08	0.12	0.11	0.09	1.00

		u										
	$x_{ij} \leq$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	$\sum$
v	1.0	0.00	0.01	0.02	0.01	0.01	0.00	0.01	0.00	0.03	0.01	0.10
	0.9	0.00	0.01	0.00	0.00	0.00	0.02	0.01	0.01	0.00	0.04	0.09
	0.8	0.00	0.00	0.00	0.02	0.03	0.02	0.00	0.02	0.02	0.00	0.11
	0.7	0.00	0.02	0.00	0.03	0.00	0.01	0.03	0.00	0.00	0.01	0.10
	0.6	0.00	0.01	0.00	0.01	0.02	0.01	0.02	0.02	0.01	0.01	0.11
	0.5	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.02	0.02	0.01	0.10
	0.4	0.00	0.01	0.04	0.01	0.00	0.01	0.01	0.02	0.01	0.00	0.11
	0.3	0.00	0.00	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.10
	0.2	0.03	0.01	0.01	0.00	0.01	0.02	0.00	0.01	0.00	0.00	0.09
	0.1	0.07	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.10
	$\sum$	0.10	0.09	0.09	0.10	0.11	0.11	0.10	0.11	0.11	0.09	1.01

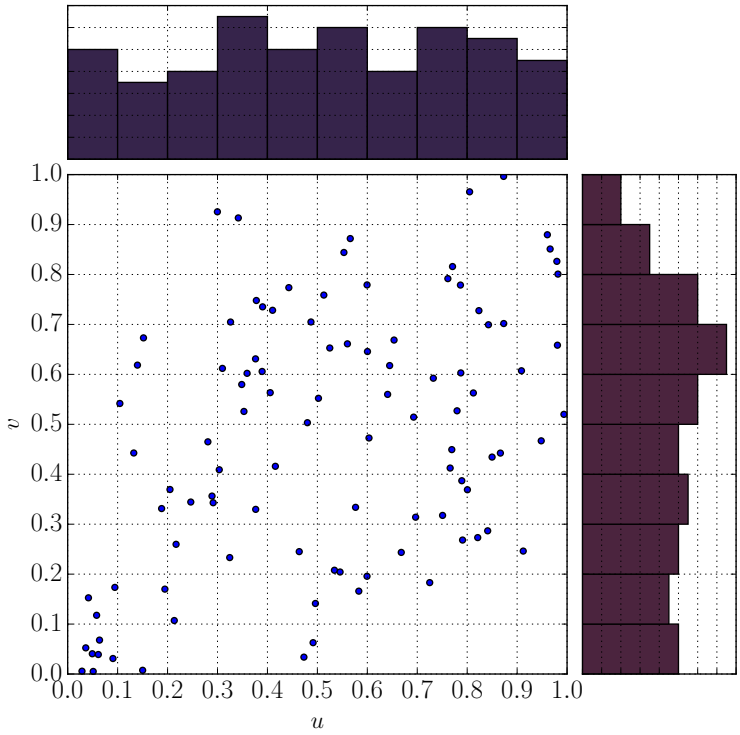

Figure 3.3.: Failure to generate uniform distribution.

with $mU_i \sim \mathcal{U}(0, m_i)$ as can be seen from:

$$\begin{aligned} C\left(\frac{v_1}{m_1}, \ldots, \frac{v_d}{m_d}\right) &= P\left(U_1 \leq \frac{v_1}{m_1}, \ldots, U_d \leq \frac{v_d}{m_d}\right) \\ &= P\left(m_1 U_1 \leq v_1, \ldots, m_d U_d \leq v_d\right) \\ &= P\left(m_1 F_1(v_1), \ldots, m_d F_d(v_d)\right) \\ &= P\left(\tilde{F}_1(v_1), \ldots, \tilde{F}_d(v_d)\right), \end{aligned}$$

where $\tilde{F}_i \equiv \mathcal{U}_i(0, m_i)$. This effect would have become stronger if we had used the empirical distribution functions of the margins $\hat{F}(u), \hat{F}(v)$ to apply the probability transformation, since the sample size is moderate and convergence to the respective uniform distributions hasn't been established at this size. Pfeifer et al., 2009 suggest imposing uniformity of the marginals in the 2-dimensional case by applying a correction to the values of the joint relative ranks in the form of the solution of

following optimization problem:

$$\min_{x_{ij}} \sum_{i=1}^{m} \sum_{j=1}^{m} (x_{ij} - a_{ij})^2 \quad \text{subject to}$$

$$\sum_{i=1}^{m} x_{ik} = \sum_{i=1}^{m} x_{lj} = \frac{1}{10} \quad \text{and} \quad x_{lk} \geq 0 \quad \text{for} \quad k, l = 1, \ldots, m.$$

The authors further suggest a two step procedure, which is computationally less expensive, by providing a general Lagrange solution to this optimization problem, which is easily solved and has to be corrected for negative values in the second step, as the Lagrange problem does not account for the non-negativity constraints. Figure 3.4 shows the solution to a 3-dimensional dataset, drawn from the product copula. We've set $m = 2$ to keep the plot simple. The probabilities $P(u = i)$, $P(v = i)$ and $P(z = 1)$ for $i = 1, 2$ are presented, they are calculated by taking the sum across the respective slices of the cube.

3.3.2. Semi-parametric estimation

Semi-parametric estimation of a copula density with the Bernstein copula density estimator can be done in three ways:

1. Specifying and estimating the parametric distributions of various subsets of the marginals, which can then be used for the probability integral transform.

2. Specifying a parametric copula and approximating it with the Bernstein copula density estimator.

3. Any combination of the former.

The usefulness of 2. becomes clear in situations, where the dependence model is known, but its estimation due to a complicated functional form is not feasible. In such a situation one could replicate this model with the Bernstein estimator since its functional form is rather simple. One would solely have to specify its components, i.e. the marginals and the copula function, and substitute this information into the Bernstein estimator.[12]
A different approach, that may be used with Bernstein polynomials and is essentially semi-parametric, is the estimation of a copula function via B-splines and a penalty to achieve a certain degree of smoothness as proposed in Kauermann et al., 2013. The general idea of estimating the density $\hat{c}_B$ as solution to a maximization problem,

[12] see Sancetta and Satchell, 2001, for an application from portfolio theory.

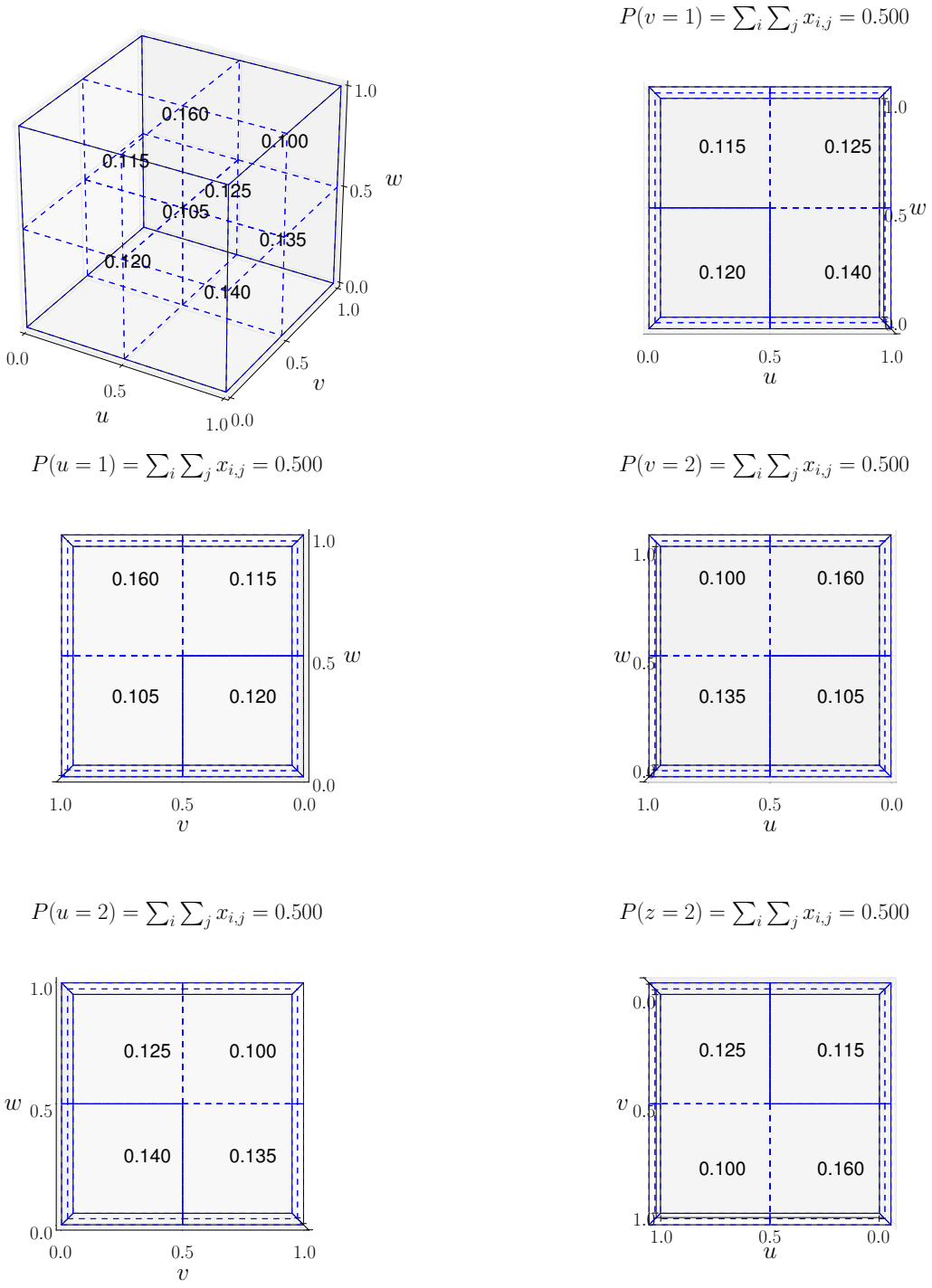

Figure 3.4.: 3d-grid, optimized to achieve uniform marginals. From top right, to bottom left the grid is rotated counterclockwise by 90°, 180°, 270°, 360°respectively. We have set $m = 2$. To assess the validity of the optimization, one can calculate the sum across the respective margins, which is 0.5 for each, thus uniformity was established.

that incorporates a penalty for smoothness was suggested by Sancetta and Satchell, 2004 who considered solving

$$\max P_n \left[\ln c_B - \lambda_n \int \left(\mathcal{D}^\alpha c_B \right)^2 \right]$$

and

$$\min P_n \left[\left(C_n - C_B \right)^2 + \int \left(\mathcal{D}^\alpha c_B \right)^2 \right],$$

where P_n and C_n denote the empirical measure and the empirical copula, λ_n is a smoothing factor that would have to be set a priori and $\mathcal{D}^\alpha$ is a differential operator of order α, i.e. $\mathcal{D}^1 = \sum_{j=1}^{k} \frac{\partial c_B}{\partial u_j}$, and $c_B = \frac{\partial^k C_B}{\partial u_1 \ldots u_k}$. Adequate constraints would have to be posed to ensure that the resulting estimate constitutes a valid copula.

The approach of Kauermann et al., 2013 is similar to these ideas and can to a certain extend be considered as a generalization thereof, as Bernstein polynomials represent one type of basis polynomials that can be applied to approximate the true copula. We will provide details to this variant of B-spline estimation after giving a short summary on the approach used by the authors. Splines are functions that are constructed by joining separate polynomial functions. The locus where two spline-segments join is called *knot* and the polynomial with the highest order within a spline specifies the degree of the spline.[13]

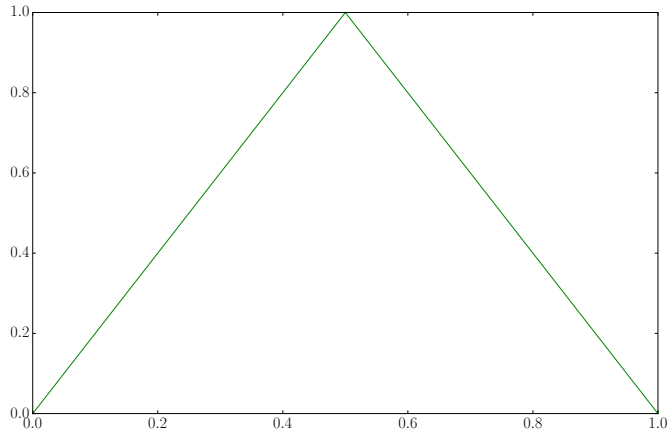

Figure 3.5.: Linear B-spline with 3 knots

Without going into the technical details we can give an intuition of the idea behind this approach as follows.[14] Multidimensional surfaces can be approximated by so

[13] see Lyche and Morken, 2008, for a succinct introduction to splines.
[14] We adapt the notation used by Kauermann et al., 2013.

called tensor product splines, which are formed by the tensor product $\boldsymbol{\phi}(u_1, \ldots, u_d) = \otimes_{j=1}^{d} \phi_j$ of the d-B-spline bases $\phi_1, \ldots, \phi_d$. The base ϕ_i represents all B-splines that can be used for approximation of the univariate density $f(u_j)$, $j = 1, \ldots, d$. The estimation of the copula density is then based on the relationship

$$c\left(u_1, \ldots, u_d\right) \approx \sum_{\mathbf{k} \in \mathcal{K}} b_{\mathbf{k}} \phi_{\mathbf{k}}\left(u_1, \ldots, u_d\right) =: c\left(u_1, \ldots, u_d; \mathbf{b}\right), \tag{3.17}$$

where $b_{\mathbf{k}}$ denotes a spline basis coefficient, for which $\sum_{k \in \mathcal{K}} b_{\mathbf{k}} = 1$ has to be imposed in the maximization step,

$$\phi_{\mathbf{k}}\left(u_1, \ldots, u_d\right) = \phi_{k_1, \ldots, k_d}\left(u_1, \ldots, u_d\right) = \prod_{j=1}^{d} \phi_{k_j}\left(u_j\right), \tag{3.18}$$

represents an element of the full tensor product $\boldsymbol{\phi}(\cdot)$ and $\phi_k(u)$ denotes a linear B-spline normalized to be a density, i.e. $\int \phi_k(u) du = 1$. By imposing the latter, as well as the requirements $c(\mathbf{u}; \mathbf{b})$ and that the estimated margins are each uniformly distributed, the authors establish that the estimate $c(\mathbf{u}, \hat{\mathbf{b}})$ is a valid copula density. The parameters that need to be estimated by maximum likelihood are the members of the set $\mathbf{b} = \{b_{\mathbf{k}}, k \in \mathcal{K}\}$, i.e. the spline basis coefficients whose number is obviously controlled by the number of elements in $\mathcal{K}$. The latter can be substantially large, when using the full tensor product for the estimation , e.g. for $K = 17$ and $d = 3$ there are $17^3 = 4913$ parameters to estimate. To reduce this number to a reasonable amount the authors suggest using a reduced form of the tensor product. We refer to Kauermann et al., 2013 for details of this approach. The log likelihood for $\mathbf{b}$ in the full tensor product setup is then given by:

$$l(\mathbf{b}) = \sum_{i=1}^{n} \log\{\sum_{k \in \mathcal{K}} b_{\mathbf{k}} \phi_{\mathbf{k}}(u_{i1}, \ldots, u_{ip}), \tag{3.19}$$

subject to the constraints mentioned above. An expression for the bias of this estimator is provided by the authors, which converges to zero exponentially in the size of the dimension of the set $\mathcal{K} = \{K_j, j = 1, \ldots, p\}$ of all individual dimensions. The variance of this estimator is not provided, but in general the behavior of B-splines is known to be similar to other non-parametric estimators in that the variance is increasing in the size of $\mathcal{K}$.[15][16]

We continue to provide the relationship between B-splines and Bernstein polyno-

[15] see Claeskens et al., 2009.
[16] see Kauermann et al., 2009.

mials. In general the jth B-spline can be expressed as

$$B_{j,m}(x) = B\left(x|t_j, \ldots, t_{j+m+1}\right),$$

where the relationship between the B-spline and the knots $(t_k)_{k=1}^{j+m+1}$ is made explicit.[17] The sequence of knots is required to be ascending and has a significant influence on the form of the spline curve. Bernstein polynomials defined on the interval $[a, b]$ can be expressed as

$$B_j^m(x) = B\left(x|\overbrace{a, \ldots, a}^{m+1-j}, \overbrace{b, \ldots, b}^{j+1}\right), \quad \text{for } j = 1, \ldots, m, \tag{3.20}$$

i.e. the sequence of knots, that have to be specified consists solely of the upper and lower bound of the interval. It can be shown that Equation 3.20 is equivalent to

$$B_j^m(x) = \binom{m}{j}\left(\frac{x-a}{b-a}\right)^j\left(\frac{b-s}{b-a}\right)^{m-j} B(x|a,b), \quad \text{for} j = 0, 1, \ldots, m, \tag{3.21}$$

which is the representation of the generalized Bernstein polynomials multiplied by the factor $B(x|a, b)$.[18] The latter is defined by

$$B(x|a,b) = \begin{cases} 1, & \text{if } a \geq x \leq b \\ 0, & \text{otherwise,} \end{cases}$$

thus ensuring, that the polynomial evaluates to 0 outside of the interval $[a, b]$. Setting $a = 0$ and $b = 1$ leads to the representation of the Bernstein copulas we introduced, except for the aforementioned factor.

$$B_j^m(x) = \binom{m}{j} x^j(1-x)^{m-j} B(x|0,1) \quad \text{for} j = 0, 1, \ldots, m, \tag{3.22}$$

Kauermann et al., 2013 conducted their estimation procedure with regular linear B-splines and B-splines with Bernstein basis polynomials as basis functions and compared the performance of these estimators against each other and a Kernel density estimate.[19]

[17] see Lyche and Morken, 2008, the authors use a slightly different notation, choosing d to denote the degree of the B-spline. Nevertheless, since the degree is equivalent to the order of the polynomial, we set $d \equiv m$ to be consistent with the notation in this thesis.

[18] In contrast to the Bernstein polynomials we introduced in this chapter, the generalized Bernstein polynomials are defined on the interval [a,b]

[19] The results seemed to suggest that the B-spline approach with regular linear B-splines performed

3.4. Alternative Representations

In this section we will present two recently developed alternative representations of the Bernstein copula, the first being based on the multivariate Baker distribution, whereas the second is based on a specific representation of the *checkerboard* copula.

3.4.1. Multivariate Baker Representation

The multivariate Baker distribution is motivated by focusing on the joint distribution of vectors of order statistics.[20] Its general form in the bivariate case is given by

$$H(x,y;R) = Pr(X_{(K)} \leq x, Y_{(L)} \leq y) = \sum_{k=1}^{m}\sum_{l=1}^{n} r_{k,l} F_{k:m}(x) G_{l:n}(y), \tag{3.23}$$

where $F_{k:m}$ and $G_{l:n}$ denote the marginal distributions of the Kth and Lth smallest order statistic generated by samples of sizes m and n of the random variables X and Y, i.e. $X_{(k)}$ and $Y_{(l)}$ respectively.[21] The matrix R consists of the entries r_{ij} which specify the probability of joint occurrence of the ranks k and l. Let $B_{k,n}(u) = \int 0^u b_{k,n}(t)dt$, $u \in [0,1]$ denote the integral over a Bernstein basis polynomial as defined in Equation (3.1). Since distribution functions of order statistics can be expressed in terms of Bernstein polynomials, one can set $F_{k:m} = mB_{k-1,m-1}(F(x))$ and $G_{l:n} = mB_{l-1,n-1}(G(y))$, thus Equation (3.23) is equivalent to

$$H(x,y;R) = C(F(x),G(y)) = mn\sum_{k=1}^{m}\sum_{l=1}^{n} r_{k,l} B_{k-1,m-1}(F(x)) B_{l-1,n-1}(G(y)), \tag{3.24}$$

which is a slightly different representation of the Bernstein copula introduced in Section 3.1. Dou et al., 2014 continue to estimate the parameters r_{ij} with the EM-algorithm, specifying m and n by minimizing the Akaike information criterion (AIC). To apply the EM-algorithm the authors specify an unobserved pair of random variables (K_i, L_i) that occur with probability $P(K_i = k, L_i = l) = r_{k,l}$, $k = 1,\ldots,m$, $l = 1,\ldots,n$ and $i = 1,\ldots,N$. The realization of this pair is linked to the indicator variable

$$\tau_{i,k,l} = \begin{cases} 1, & \text{if } (K_i, L_i) = (k,l) \\ 0, & \text{otherwise,} \end{cases} \tag{3.25}$$

best, nevertheless while the dimension of the Bernstein polynomial basis was chosen by comparisons of the corrected Akaike Information Criterion as the interval $[1,\ldots,10]$, it is not clear to us which degree m was specified for the Bernstein basis polynomials.

[20] see Baker, 2008, the author introduces this distribution for the bivariate case, nevertheless the extension to the multivariate case is straightforward.

[21] see Dou et al., 2014.

the expectation of which, conditional on a realization of (x_i, y_i),

$$\hat{\tau}_{i,k,l} = \mathrm{E}\left[\tau_{i,k,l} | (x_i, y_i)_{1 \leq i \leq N}; R\right] \tag{3.26}$$

is calculated in the E-step of the EM algorithm. The likelihood of this model in the case of two continuous marginals is then given by

$$\prod_{i}^{N} \prod_{k}^{m} \prod_{l}^{n} \left[r_{k,l} f_{k:m}(x_i) g_{l:n}(y_i)\right]^{\tau_{i,k,l}}. \tag{3.27}$$

The authors apply this methodology to various datasets and note that the AIC has a tendency to prefer smaller numbers for the dimension of R, i.e. in our terminology for specifying the grid size m. In general we would suggest to use the corrected Akaike information criterion $AICc$ to account for the well known property of the AIC to prefer models with too many parameters.[22]

3.4.2. Rook Copula

Cottin and Pfeifer, 2014 introduced so called rook copulas which can be viewed as *checkerboard* copulas with fixed grid size $m = m_i,\ i = 1, \ldots, d$ for all dimensions. The benefit of this representation is that it reduces the computational effort of evaluating the multiple sums in 3.8 when calculating the Bernstein and *checkerboard* copula from $\mathcal{O}\left(m^d\right)$ to $\mathcal{O}\left(m \cdot d\right)$. To see this we have to look at the definition of a rook copula:[23]

$$p_m\left(\nu_1, \ldots, \nu_d\right) = P\left(\bigcap_{i=1}^{d} \{U_i = \nu_i\}\right) = \frac{1}{m} \Leftrightarrow \left(\nu_1, \ldots, \nu_d\right) = \left(\sigma_{t,1}, \ldots, \sigma_{t,d}\right), \tag{3.28}$$

where the vector $(\sigma_{i,1}, \ldots, \sigma_{i,d},\ i = 1, \ldots, n)$ indicates the rank of each variable in sample i. For each of the d- variables we can define the column vectors $(\sigma_{0,d}, \ldots, \sigma_{m-1,d})$, $m = n$ for every dimension and collect them in the permutation matrix:

$$M := \begin{pmatrix} \sigma_{0,1} & \sigma_{0,2} & \ldots & \sigma_{0,d-1} & \sigma_{0,d} \\ \sigma_{1,1} & \sigma_{1,2} & \ldots & \sigma_{1,d-1} & \sigma_{1,d} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{m-2,1} & \sigma_{m-2,2} & \ldots & \sigma_{m-2,d-1} & \sigma_{m-2,d} \\ \sigma_{m-1,1} & \sigma_{m-1,2} & \ldots & \sigma_{m-1,d-1} & \sigma_{m-1,d}. \end{pmatrix}$$

[22] see e.g. Burnham and Anderson, 2004.
[23] see Cottin and Pfeifer, 2014.

Thus the locus of all points in the d-dimensional hypercube are specified by $m \cdot d$ coordinates. The drawback of this construction is clearly that the grid size m is always as large as the sample size n, which from our experience leads to overfitting. The right graphic in Figure 3.6 depicts a rook copula that has been constructed from 25 samples from a bivariate Clayton copula

$$C_{Cl} = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}, \tag{3.29}$$

with parameter $\theta = 3$, which is shown in the left graphic of the same figure. Being a *checkerboard*-copula the rook copula is not smooth, but one can easily construct samples from the corresponding Bernstein copula using the same methodology. We refer to the authors for details regarding the simulation from rook copulas.

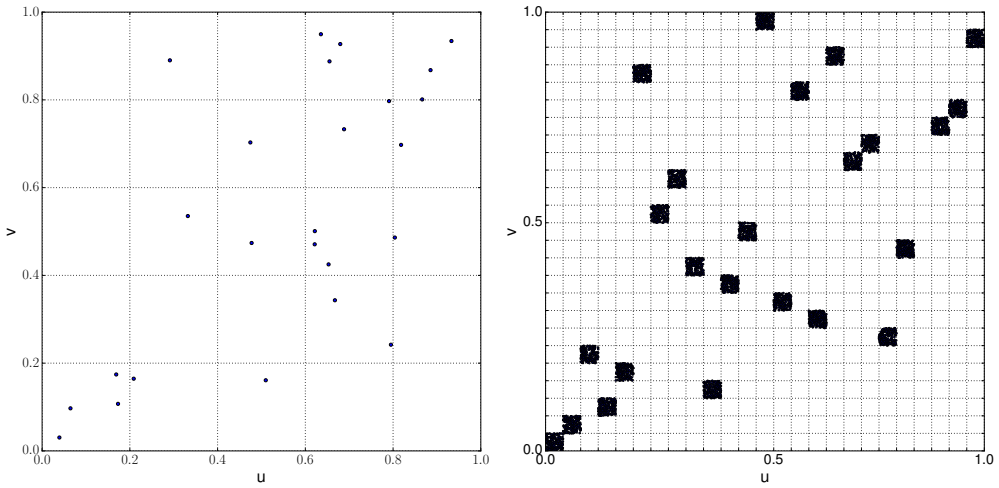

Figure 3.6.: Left figure: Original sample - Clayton copula ($\theta = 3$).
Right figure: 5000 samples from a rook copula.

3.5. Application: Model selection

In the following we present a simulation study with the purpose to assess the precision of the Bernstein copula density estimator for samples of medium size. The large sample properties of this estimators seem promising, nevertheless in real world applications one is often confronted with situations where data is scarce. We are interested in a real world scenario, in which the analyst has no information with regards to the true underlying probability model, thus she has to select a proper model from various candidate models. We draw samples from the bivariate Clayton-,

Frank-, Gauss-, and t-copulas $(C_\theta^{Cl}, C_\theta^{Fr}, C_\rho^{Ga}, C_{\nu,\rho}^t)$ and estimate the copula densities for each of these samples from the set $\{\hat{c}_\theta^{Cl}, \hat{c}_\theta^{Fr}, \hat{c}_\theta^{Ga}, \hat{c}_\theta^t, \hat{c}_B\}$, where $\hat{c}_B$ denotes the Bernstein copula density estimator. The parameter estimates for the parametric copulas are obtained by Maximum likelihood. In Figure 3.7 we have generated scatter plots from the 4 copula models according to the parameterization given in the plot titles. For C_ρ^{Ga} and $C_{\nu,\rho}^t$ we have additionally set the marginal distributions to the standard normal and standard Student's t distributions respectively.

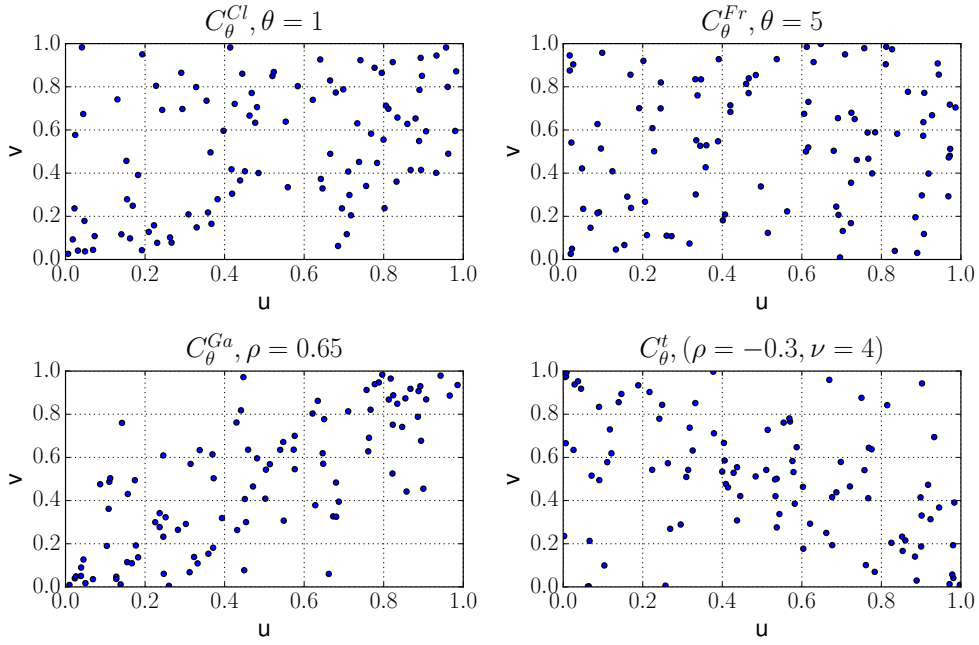

Figure 3.7.: Model selection: Scatter plots of 100 samples from the true copula models.

The respective copulas ot these models in the bivariate case are given by[24]

$$
\begin{aligned}
C_\theta^{Fr}(u_1, u_2) &= -\frac{1}{\theta} \ln\left(1 + \frac{(exp\,(-\theta u_1) - 1)\,(exp\,(-\theta u_2)) - 1}{exp\,(-\theta) - 1}\right), \\
C_\rho^{Ga}(u_1, u_2) &= \mathbf{\Phi}_\rho\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right), \\
C_{\nu,\rho}^t(u_1, u_2) &= \mathbf{t}_{\nu,\rho}\left(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)\right).
\end{aligned}
$$

The commonly used metrics to select between various parametric models are the AIC and the Bayesian information criterion BIC.[25] Since both of these criteria

[24] Regarding the formula for the Clayton copula see Equation (3.29).

[25] Even though it is recommended to use $AICc$ instead of AIC to account for AIC's overfitting bias in small to medium samples, AIC is still widely used in scientific literature.

are not feasible for non-parametric estimators we have to apply a different metric to select between the Bernstein copula density estimator and a set of parametric models. In this regard we choose to select according to the relative distance ($RD_{N,p}$) given by

$$RD_{N,p} = \frac{\|c_\theta - c_e\|_{N,p}}{\|c_e\|_{N,p}}, \quad p \in \{1, 2, \infty\}, \tag{3.30}$$

where N denotes the sample size. We thus select according to the p-norm of the distance between the true copula density c_θ and the estimate c_e. Table 3.2 summarizes the means of the parameter estimates for the unknown copula densities. The parameterization was chosen with the intention to cover a wide variety of dependence structures. Note that the negative estimate for c_θ^{Cl}, which is a result of maximum likelihood trying to capture the negative linear dependence structure of the specified t-copula, leads to a non-strict copula.[26] Table 3.3 depicts the mean of

Table 3.2.: Parameter estimates for the model-selection simulation

	c_θ^{Cl}	c_θ^{Fr}	c_θ^{t}	c_θ^{Ga}
$\bar{\theta}$	$\theta = 1$	$\theta = 5$	($\rho = -0.3$, $\nu = 4$)	$\rho = 0.65$
$\hat{c}_\theta^{Cl}$	0.80603	0.17547	-0.06504	1.37974
$\hat{c}_\theta^{Fr}$	2.72932	0.68353	-3.41627	6.52156
$\hat{c}_\theta^{Ga}$	0.44453	0.10267	-0.38969	0.74300
$\hat{c}_\theta^{t}$	(0.44128, 4.1719)	(0.10413, 54.93915)	(-0.49617, 2.7126)	(0.74579, 13.06077)

the $AICc$- and BIC-values across the 100 samples. Only for the Clayton copula we observe that the correctly specified model attains the minimum $AICc$- and BIC-value. For the rest of the samples a misspecified copula is selected to be the best model on average. The obvious reason for this is that at a medium sample size of 100 the characteristics of the sample at hand do not convey enough information to adequately discriminate between true and false models.

[26] see e.g. Nelsen (2007), an Archimedian copula is considered non-strict when its generator $\phi(0) \neq \infty$. To avoid undefined areas in a non strict Archimedian copulas one has to resign to the notion of a pseudo inverse, when constructing these copulas.

Table 3.3.: Mean of $AICc$ and BIC values from 100 simulations with sample size 100 each. Parameters for the true copula densities: $(\hat{c}_\theta^{Cl}, \theta = 1)$,$(\hat{c}_\theta^{Fr}, \theta = 5)$,$(\hat{c}_\theta^{Ga}, \rho = 0.65)$, $(\hat{c}_\theta^{t}, (\rho = -0.3, \nu = 4)$

		True			
	$\overline{AICc}$	c_θ^{Cl}	c_θ^{Fr}	c_θ^{Ga}	c_θ^{t}
Estimate	$\hat{c}_\theta^{Cl}$	-25.08766	-0.14396	0.05574	-62.37794
	$\hat{c}_\theta^{Fr}$	-17.11867	0.51655	-23.78354	-77.12786
	$\hat{c}_\theta^{Ga}$	-16.41790	2.86301	-21.48877	-76.74386
	$\hat{c}_\theta^{t}$	-18.22732	5.03348	-46.59113	-75.20874
	$\overline{BIC}$	$\hat{c}_\theta^{Cl}$	$\hat{c}_\theta^{Fr}$	$\hat{c}_\theta^{Ga}$	$\hat{c}_\theta^{t}$
Estimate	$\hat{c}_\theta^{Cl}$	-22.5233	2.42039	2.62009	-59.81359
	$\hat{c}_\theta^{Fr}$	-14.55432	3.08090	-21.21919	-74.56350
	$\hat{c}_\theta^{Ga}$	-11.33127	7.94964	-16.40214	-71.65723
	$\hat{c}_\theta^{t}$	-10.66181	12.59899	-39.02562	-67.64323

In Table 3.4 we compare all estimators, including the Bernstein copula density estimator $\hat{c}_B$ according to their $RD_{100,1}$-values. To obtain these values we have randomly chosen 1000 points from $]0,1[^d$ and computed this metric for each of the estimators at each of the 100 samples. We can observe that $\hat{c}_B$ performs reasonably well on average, minimizing $RD_{100,1}$ at the estimation of the Clayton- and Frank copulas, while generating the second best $RD_{100,1}$-values for the Gauss- and t-copulas. From the box plots in Figure 3.8 we can see that for the estimation of the Clayton- and Frank copula densities the relative distance of all estimators differ only slightly. This result may be explained by reevaluating the scatter plots in Figure 3.7, from which we can see that the distribution of the samples in the plane for these copulas is rather homogeneous, thus revealing only minor information to the specific structure of the data generating process. If we choose $RD_{100,2}$ and $RD_{100,\infty}$ the results are clearly different, since these metrics are functionals of norms that punish large deviations stronger. Since $\hat{c}_B$ is a non-parametric estimator it has rather poor performance in the estimation of the tails of the copula density, for which $c(\mathbf{u}) \to \infty, \mathbf{u} = (u_1, \ldots, u_d)$ holds. Parametric estimators do not suffer from this drawback thus deviations in the tail, that result from misspecification, are in general of lesser extent. The results of the $RD_{100,2}$ and $RD_{100,\infty}$ evaluations are reported in Section A.1 of the Appendix.

The results of this simulation study suggest that the Bernstein copula density estimator performs well, providing robust estimates even if it is compared to parametric estimators that can cover a wide range of dependence structures such as the

Table 3.4.: Mean of $RD_{100,1}$-values from 100 simulations with sample size 100 each. 95% Confidence Intervals are reported in brackets. The minimum values are highlighted blue. Parameters for the true copula densities: $(\hat{c}_\theta^{Cl}, \theta = 1)$,$(\hat{c}_\theta^{Fr}, \theta = 5)$,$(\hat{c}_\theta^{Ga}, \rho = -0.3)$, $(\hat{c}_\theta^{t}, (\rho = 0.65, \nu = 4)$

RD	$\hat{c}_\theta^{Cl}$	$\hat{c}_\theta^{Fr}$
$\hat{c}_B$	0.71103, (0.70975, 0.7123)	0.69730, (0.6959, 0.69873)
$\hat{c}_\theta^{Cl}$	0.71142, (0.70982, 0.71331)	0.69736, (0.69597, 0.69875)
$\hat{c}_\theta^{Fr}$	0.71110, (0.7098, 0.71236)	0.69732, (0.69598, 0.69867)
$\hat{c}_\theta^{t}$	0.71608, (0.71461, 0.71775)	0.69736, (0.69602, 0.69868)
$\hat{c}_\theta^{Ga}$	0.71265, (0.7113, 0.71394)	0.69736, (0.69599, 0.69868)
	$\hat{c}_\theta^{Ga}$	$\hat{c}_\theta^{t}$
$\hat{c}_B$	0.70133, (0.69993, 0.70267)	0.80212, (0.80095, 0.80332)
$\hat{c}_\theta^{Cl}$	0.69917, (0.69716, 0.70178)	0.77425, (0.77284, 0.77564)
$\hat{c}_\theta^{Fr}$	0.74648, (0.74493, 0.74796)	0.85094, (0.84977, 0.85215)
$\hat{c}_\theta^{t}$	0.75363, (0.75198, 0.75549)	0.82781, (0.82601, 0.82977)
$\hat{c}_\theta^{Ga}$	0.75778, (0.7562, 0.75954)	0.82373, (0.82241, 0.82499)

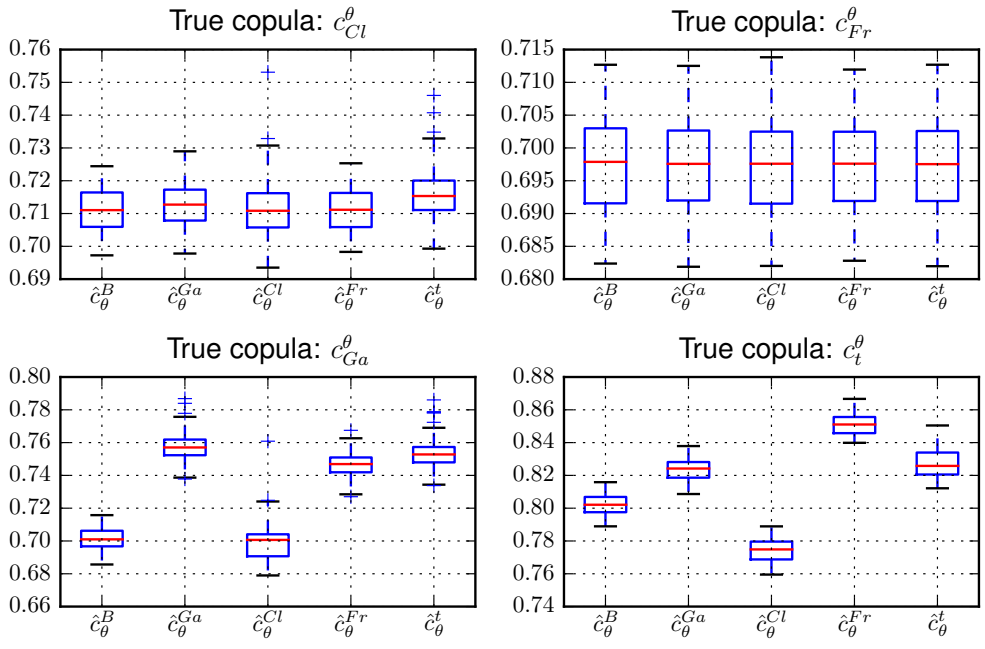

Figure 3.8.: Model selection: Box plots for mean values of $R_{100,1}$-values.

Gauss copula or the t copula. This comes with the benefit of not having to specify any assumptions with regard to the data generating process. Nevertheless, for applications that require good estimates of regions in the outer tail of the distribution a flexible parametric model might be a more sensible choice.[27]

[27] These results are of course obtained in a very low dimensional setting, see Weiss and Scheffer, 2012, for an analysis of the performance of density estimatiors based on Bernstein polynomials.

4. Bandwidth Selection

In this chapter we consider the problem of choosing an optimal bandwidth parameter m for the Bernstein copula, and propose an estimator that is based on the local rank correlation of the respective sample we wish to analyse. In the majority of the present literature that deals with the estimation of copulas and copula densities via Bernstein polynomials the bandwidth parameter is specified either ad-hoc or via grid searches against a specific loss function. An exception is made in the case of rook copulas, for which m is identical to the sample size by definition. Bouezmarni et al., 2013 introduce a cross validation procedure to arrive at a theoretically sound estimate of m, which we will shortly discuss in the next section, and which we will utilize as benchmark for the estimator we develop in the Sections 4.2 and 4.3.

4.1. Cross Validation

The least-squared cross validation $LSCV$ method for estimating an optimal grid size m leads to an estimator $\hat{m}$ that minimizes

$$LSCV(m) = \int_0^1 \hat{c}^2(u,v)dudv - 2n^{-1}\sum_{i=1}^{n} \hat{c}^{(-i)}\left(X_{1,i}, X_{2,i}\right). \tag{4.1}$$

It can be shown that by taking the expectation of Equation (4.1) one arrives at

$$\mathrm{E}\left(LSCV(m)\right) = \mathrm{E}\left(\int_0^1 \left(\hat{c}\left(u,v\right) - c\left(u,v\right)\right)^2 dudv\right) - \int_0^1 c^2\left(u,v\right)dudv, \tag{4.2}$$

thus minimizing Equation (4.1) is equivalent to minimizing an unbiased estimator of the expected integrated square error (ISE).[1] This type of bandwidth estimator has been known to provide rather volatile results within Kernel density estimation due to slow rates of convergence at $\mathcal{O}(n^{\frac{1}{10}})$. In particular it can be shown that

$$n^{\frac{1}{10}}\left(\frac{h_{LSCV}}{\hat{h}_{ISE}} - 1\right) \sim \mathcal{N}(0, \sigma^2_{LSCV}), \tag{4.3}$$

[1] see Bouezmarni et al., 2013.

where h_{LSCV} denotes the least-squared cross validation bandwidth and $\hat{h}_{ISE}$ denotes the bandwidth that minimizes the ISE of the density estimate $\hat{f}_h$, $ISE(\hat{f}_h)$.[2] There is an alternative to $LSCV(h)$ given by the biased cross-validation estimator $BCV(h)$ with identical rate of convergence, but with variance σ_{BCV} that relates to σ_{LSCV} as $\frac{\sigma_{BCV}}{\sigma_{LSCV}} \approx 15.7$. Developing an estimator of this kind should be feasible for estimating m as well, and is part of future research. see Bouezmarni et al., 2013 state that

$$\int_0^1 \hat{c}^2(u,v)dudv = \frac{1}{n^2}\sum_i \sum_j \tilde{B}\left(\lfloor X_{1,i}m \rfloor, \lfloor X_{1,j}m \rfloor\right) \tilde{B}\left(\lfloor X_{2,i}m \rfloor, \lfloor X_{2,j}m \rfloor\right), \quad (4.4)$$

whith

$$\tilde{B}\left(a,b\right) = \frac{B\left(a+b+1, 2m-a-b-1\right)}{B\left(a+1, m-a\right) B\left(b+1, m-b\right)},$$

where $B\left(a,b\right) = \int_0^1 t^{a-1}(1-t)^{b-1}$ denotes the Beta function.

4.2. Mean Squared Error Optimality

Our goal is to specify the grid size m in a manner that can be considered optimal with respect to a suitable loss function. The bandwidth h in kernel density estimation fulfils a similar role to m as it specifies the degree of smoothness that is applied to a given dataset. If the bandwidth is chosen too small the estimated density is *undersmoothed* and appears to be "rough", as too much weight is given to the individual data points. A bandwidth that is too large *oversmoothes* the data and the information provided by the dataset is not taken to account to an adequate extent. To specify an optimal bandwidth it is common practice in kernel density estimation to choose that h that is optimal with respect to the asymptotic mean integrated squared error (AMISE), which is the asymptotic counterpart to the mean integrated squared error given by:

$$\text{MISE}(h) = \text{E}\int (\hat{f}_h(x) - f(x))^2\, dx.$$

The resulting expression for h suffers from the drawback, that it is still dependent on functionals of f. This leads to various possibilities of specifiying the unknown functionals and there has been a large amount of research with regards to this topic.[3] A similar approach to specifying an optimal grid size m for the Bernstein copula density c_B is not feasible as its functional form, i.e. the number of summands in its

[2] see Sheather, 2004.

[3] For an overview, see e.g. Sheather, 2004, The author classifies the different approaches into rules of thumb, cross-validation methods and so called plug - in estimators.

representation is dependent on m, thus c_B is not specified unless m has been set in advance. Therefore we propose to take a different, data-driven, approach to arrive at an optimal grid size m_{opt}. In the following we will present a few facts on our chosen loss function, the mean squared error (MSE) and then develop a criterion that allows us to specify an optimal grid size based on its rate of increase when optimized in a mean squared error sense.
The mean squared error of a copula density estimator is defined as:

$$\begin{aligned} \mathrm{MSE}\left[\hat{c}\left(\mathbf{u}\right)\right] &= \mathrm{E}\left[\hat{c}\left(\mathbf{u}\right)\right] - c\left(\mathbf{u}\right) \\ &= \mathrm{Bias}\left[\hat{c}(\mathbf{u})\right]^2 + \mathrm{Var}\left[\hat{c}(\mathbf{u})\right], \end{aligned} \tag{4.5}$$

where $\mathrm{Bias}\left[\hat{c}\left(\mathbf{u}\right)\right]$ is defined as:

$$\mathrm{Bias}\left[\hat{c}(\mathbf{u})\right] = \mathrm{E}\left[\hat{c}\left(\mathbf{u}\right)\right] - c\left(\mathbf{u}\right),$$

and $\mathrm{Var}(\hat{c})$ denotes the variance of the estimator:

$$\mathrm{Var}\left[\hat{c}(\mathbf{u})\right] = \mathrm{E}\left[\hat{c}(\mathbf{u}) - \mu_{\hat{c}(\mathbf{u})}\right]^2.$$

Non parametric density and copula estimators are known to face a bias-variance trade off with respect to the mean squared error, i.e. adjusting the bandwidth parameter according to the limiting behavior that is necessary for convergence to the true underlying estimand leads to a decrease in the bias, but simultaneously to an increase in the variance of the estimator. This is the basis on which optimizations with respect to the bandwidth parameter can be conducted. Since the $\mathrm{MSE}\left[\hat{c}(\mathbf{u}\right]$ is convex in $\mathrm{E}\left[\hat{c}(\mathbf{u})\right]$ there is an optimal bandwidth to be derived from its integrated counterpart the $\mathrm{MISE}\left[\hat{c}(\mathbf{u})\right]$ and procedures as mentioned above are feasible. This follows from analogy to the bandwidth selection problem in kernel density estimation.

Let m_{opt} denote the optimal grid size according to the mean squared error criterion:

$$m_{opt} = \arg\min_{m} \mathrm{MSE}\,. \tag{4.6}$$

The following result was established by Sancetta and Satchell (2004):

$$m_{opt} = \mathcal{O}\left(n^{\frac{2}{k+4}}\right). \tag{4.7}$$

Equation (4.7) specifies the rate at which m_{opt} increases with respect to the sample size n.

Let G denote the set of functions with asymptotic behavior as specified by (4.7):

$$G = \left\{ f : \frac{\delta f}{\delta n} = O(n^{\frac{2}{k+4}}) \right\}. \tag{4.8}$$

Elements of G are possible specifications of a relationship between m_{opt} and n that could be exploited to estimate m_{opt}.
Thus the we are seeking to choose that element of G that yields function values that are as close as possible to the theoretical optimum (4.6):

$$m_{opt} = \underset{g(n)}{\arg\min}\ \mathrm{MSE}\,. \tag{4.9}$$

A linear solution for $g(n)$ in (4.9) would be of the form:

$$m_{opt} = a + b(n^{\frac{2}{k+\mathrm{Bias}\,4}}), \tag{4.10}$$

Until know the discussion is still very general and we can clearly see from (4.8), that the correct specification of g is non-trivial, as there are infinitely many elements in G. We propose to utilize information provided by the dataset in the form of a dependence measure θ to establish a natural relationship between the sample size in the optimal growth rate (4.7) and m_{opt}.[4] An advantage of this specification is, that it enables us to treat $g(n, \theta)$ as an estimator for m_{opt}, since θ is a probabilistic quantity. For the linear specification (4.10) one could then set:

$$m_{opt} = |\theta|(n^{\frac{2}{k+4}}). \tag{4.11}$$

Since $m > 2$ has to hold it is reasonable in this case to set $b = 0$. (4.11) implies that stronger dependency in a given dataset requires less smoothing i.e. a higher m, since m_{opt} is increasing in θ. An intuition for the latter might be gained by considering the relationship between concordance measures and copulas. Let θ to be one of the popular concordance measures, i.e. Kendall's tau or Spearman's rho. Since both of these statistics are functionals of a bivariate copula, they inherit its property of being invariant under strictly increasing component-wise transformations[5]. They thus can be considered as copula properties, giving a natural relationship between the dependence structure of a random vector and a corresponding dependence measure. A high $|\theta|$ in a reasonably sized random sample indicates a distinctive dependence structure that can be captured more appropriately by applying a finer grid. In contrast $|\theta| \to 0$ the copula approaches the independence copula, where data points are distributed rather symmetrically in the hypercube. In this scenario a low m is

[4] Bouezmarni et al., 2013, indicate that such a specification might be reasonable.
[5] see e.g. McNeil et al., 2010.

appropriate since it captures the symmetrical nature of the dependence structure more robustly. Figure 4.1 shows possible curves that are fitted against simulated

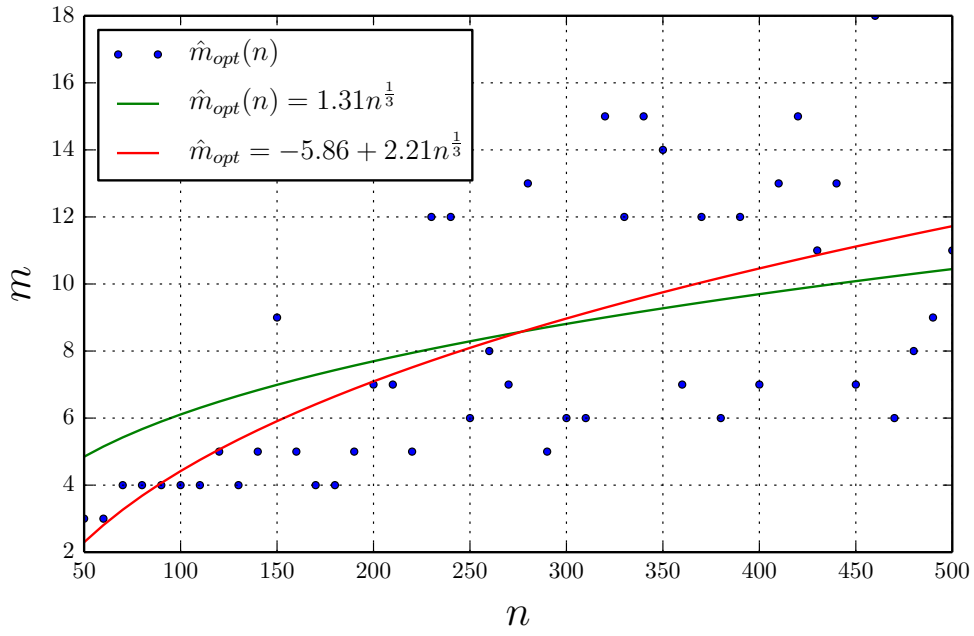

Figure 4.1.: Optimal grid sizes with linear specifications - Data are MISE-optimal grid sizes for the 2 dimensional Gauss Copula: $C^{ga}_{\rho=0.9}(u)$.

estimates of MISE- optimal grid sizes. Both curves were created by fitting linear models to the data. Since we can't observe the estimates of the MSE-optimal grid sizes directly a regression approach isn't possible in reality. We therefore have to develop a model from which we can derive these estimates.

4.3. Penalized Grid Optimality

We allow m to take values in $\mathbb{R}^+$ and specify the following criterion:[6]

$$\hat{m}_{\mathsf{p}} = \arg\max_{m} f_p$$
$$where\, f_p = \frac{n}{m}\left(\frac{m}{n}\right)^{1-|\theta|^{-\frac{1}{n}}} \ln\left(\frac{1}{|\theta|} m n^{-\frac{2}{d+4}}\right), \tag{4.12}$$

[6]We relax the integer value restriction for the purpose of the analysis. When specifiying m in applications we will choose $m = nint(\hat{m}_{\mathsf{p}})$, where $nint(x)$ denotes the nearest integer function.

Equation (4.12) specifies a criterion which allows us to operationalize the selection of m. In this respect this optimization problem may be viewed as selecting the optimal c_B^m from the family $\mathcal{B}$ given by:

$$\mathcal{B} = \{c_B^m,\, m \in \{2, 3, \dots\}\}.$$

We can regard $\mathcal{B}$ as consisting of a single Bernstein-copula-density estimator at all stages of convergence towards the true underlying density c. We further assume that the optimal grid size is dependent on the level of pairwise dependency within the random vector in consideration, see above. Finally, we require $\theta \in [-1, 1]$ and

$$\theta\,(u, v) = -1 \iff W(u, v) = \max(u + v - 1, 0) \tag{4.13}$$
$$\theta\,(u, v) = 0 \iff \prod(u, v) = uv \tag{4.14}$$
$$\theta\,(u, v) = 1 \iff M(u, v) = \min(u, v). \tag{4.15}$$

For $\theta = 1$, Equation (4.12) then reduces to:

$$f_p = \frac{n}{m} \ln\left(mn^{-\frac{2}{d+4}}\right),$$

whereas for $\theta = 0$, f_p is not defined. For $\theta \in]0, 1[$ we can set

$$\begin{aligned}\alpha &= 1 - |\theta|^{-\frac{1}{n}} \\ &= 1 - \frac{1}{|\theta|^{\frac{1}{n}}} < 0,\end{aligned}$$

so that the second factor in 4.12 becomes

$$\left(\frac{m}{n}\right)^{1-|\theta|^{-\frac{1}{n}}} = \left(\frac{n}{m}\right)^{-\alpha}$$

and (4.12) can be expressed as:

$$\begin{aligned}f_p &= \left(\frac{n}{m}\right)^{1-\alpha} \ln\left(\frac{1}{|\theta|} mn^{-\frac{2}{d+4}}\right) \\ &= \left(\frac{n}{m}\right)^{1-\alpha} \left[\ln\left(\frac{1}{|\theta|}\right) + \ln\left(mn^{-\frac{2}{d+4}}.\right)\right] \\ &= \left(\ln mn^{-\frac{2}{d+4}}\right) \left(\frac{n}{m}\right)^{1-\alpha} + \mathcal{O}\left(\frac{1}{m^{1-\alpha}}\right)\end{aligned} \tag{4.16}$$

Since we will be optimizing with respect to m we can treat θ and n as constant and focus on the behavior of f_p with respect to m. From (4.16) we can see that f_p is

essentially controlled by the expression

$$\ln\left(mn^{-\frac{2}{d+4}}\right)\left(\frac{n}{m}\right)^{1-\alpha}, \tag{4.17}$$

and that increasing values of m lead to increasing values of this criterion through the first factor of (4.17), but are simultaneously punished by the second factor. We specify this as it seems reasonable, when comparing multiple $\hat{m}$ that satisfy (4.7), to prefer that estimator, that relies on the highest ratio of data points per grid. The first factor in expression (4.17) is composed of the product of m and the optimal rate of increase (4.7). The latter is constant in our optimization problem (4.12), thus it will be part of the optimal value obtained by the maximization, ensuring that the latter increases accordingly in n.
We see that f_p is continuous and differentiable everywhere except at $f_p(0)$, where it is continuous but not differentiable due to the use of the absolute value of the parameter θ . We chose the latter to induce symmetry with respect to the direction of the dependence captured by this measure.

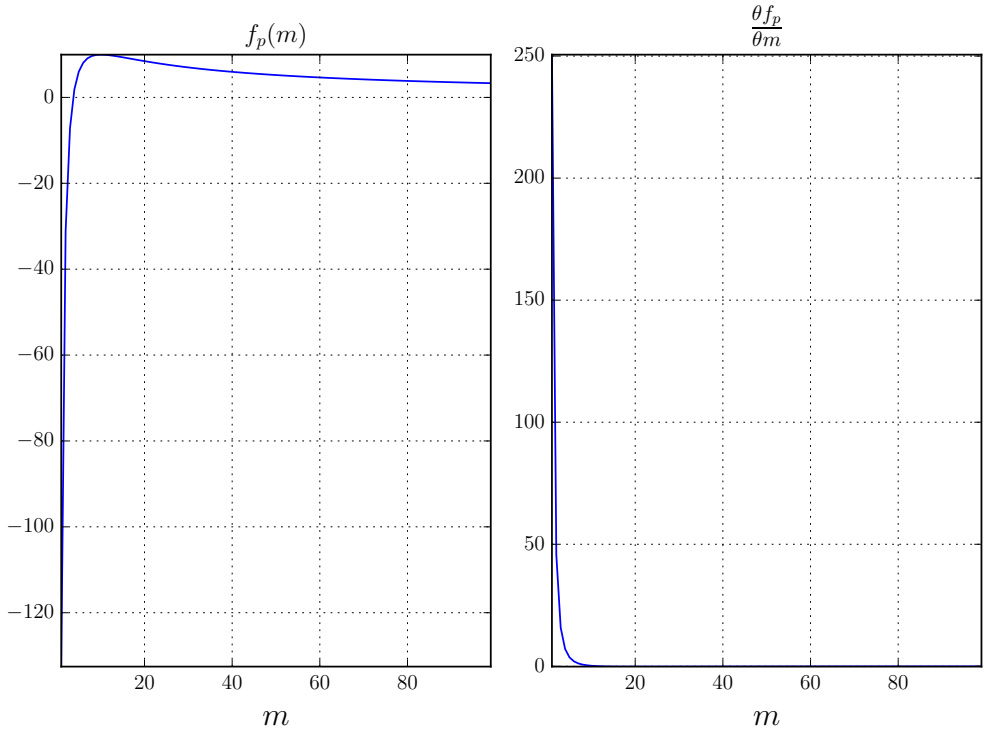

Figure 4.2.: Plot of f_p and $\frac{\theta f_p}{\theta m}$, for $n = 100$, $d = 2$, $\theta = 0.8$.

f_p obtains a maximum at:

$$\hat{m}_{\mathsf{p}} = |\theta| e^{|\theta|^{\frac{1}{n}}} n^{\frac{2}{d+4}} \tag{4.18}$$

thus satisfying (4.7). (4.18) is a global optimum, if

$$\begin{aligned}\frac{d^2 f_p}{dm^2}(\theta) = \frac{|\theta|^{-\frac{2}{n}}}{m^2}\left(\frac{m}{n}\right)^{-|\theta|^{-\frac{1}{n}}}\Bigg(ln\left(\frac{m}{|\theta|}n^{-\frac{2}{d+4}}\right)|\theta|^{\frac{1}{n}} \\ + \ln\left(\frac{m}{|\theta|}n^{-\frac{2}{d+4}}\right) - |\theta|^{\frac{2}{n}} - 2|\theta|^{\frac{1}{n}}\Bigg) < 0.\end{aligned} \tag{4.19}$$

As $m, n, d > 0$, (4.19) will only hold for a suitable value of θ. To be precise, we are interested in the set $D = \{\theta : \frac{d^2 f_p}{dm^2}(\theta) < 0\}$. Let $g(\theta) = \frac{d^2 f_p}{dm^2}(\theta)$. $g(\theta)$ is quadratic in the absolute value of θ, $|\theta|$. As $|\theta|$ is not differentiable on its complete support a piece-wise analysis of the properties of $g(\theta)$ is necessary to specify which θ are elements of D . We separate the support of g at $\theta = 0$ and denote $D_1 = \{\theta : \theta < 0\}$ and $D_2 = \{\theta : \theta > 0\}$. Using the definition of the absolute value:

$$|x| = \begin{cases} x, & \text{if } x \geq 0 \\ -x, & \text{if } x < 0, \end{cases} \tag{4.20}$$

we define the functions $g_1 : D_1 \to \mathbb{R}$ and $g_2 : D_2 \to \mathbb{R}$:

$$\begin{aligned}g_1 = \frac{(-\theta)^{-\frac{2}{n}}}{m^2}\left(\frac{m}{n}\right)^{-(-\theta)^{-\frac{1}{n}}}\Big[-(-\theta)^{\frac{2}{n}} + (-\theta)^{\frac{1}{n}}\ln\left(-\frac{m}{\theta}n^{-\frac{2}{d+4}}\right) - 2(-\theta)^{\frac{1}{n}} \\ + \ln\left(-\frac{m}{\theta}n^{-\frac{2}{d+4}}\right)\Big],\end{aligned} \tag{4.21}$$

$$g_2 = \frac{\theta^{-\frac{2}{n}}}{m^2}\left(\frac{m}{n}\right)^{-\theta^{-\frac{1}{n}}}\left[-\theta^{\frac{2}{n}} + \theta^{\frac{1}{n}}\ln\left(\frac{m}{\theta}n^{-\frac{2}{d+4}}\right) - 2\theta^{\frac{1}{n}} + \ln\left(\frac{m}{\theta}n^{-\frac{2}{d+4}}\right)\right]. \tag{4.22}$$

Equations (4.21) and (4.22) are not trivially solved, thus we conduct a case analysis on (4.21) to investigate necessary conditions for them to be negative. Due to the symmetry of the absolute value the same logic will hold for necessary conditions on (4.22). Since the first factor in g_1 is always positive, we will focus on the second factor and set:

$$\begin{aligned}a &= (-\theta)^{\frac{1}{n}}\ln\left(-\frac{m}{\theta}n^{-\frac{2}{d+4}}\right) + \ln\left(-\frac{m}{\theta}n^{-\frac{2}{d+4}}\right) \\ &= \ln\left(-\frac{m}{\theta}n^{-\frac{2}{d+4}}\right)\left((-\theta)^{\frac{1}{n}} + 1\right), \\ b &= -(-\theta)^{\frac{2}{n}} - 2(-\theta)^{\frac{1}{n}}.\end{aligned}$$

We observe that:

$$a > 0 \iff \ln\left(-\frac{m}{\theta}n^{-\frac{2}{d+4}}\right) > 0, \tag{4.23}$$

$$a < 0 \iff \ln\left(-\frac{m}{\theta}n^{-\frac{2}{d+4}}\right) < 0. \tag{4.24}$$

Thus if (4.23) holds, we have:

$$a + b < 0 \iff |a| < |b|,$$
$$a + b > 0 \iff |a| > |b|.$$

If (4.24) holds $a+b$ is always negative. Thus if the sample size is sufficiently larger than the number of grids, $n >> m$, (4.24) will hold and (4.18) is a global optimum.

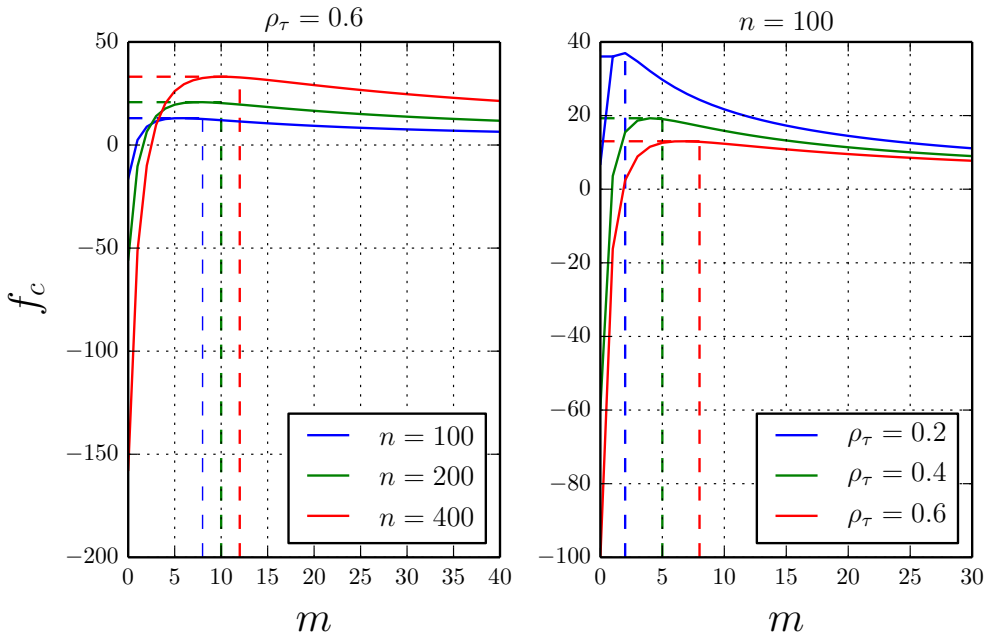

Figure 4.3.: f_p in the 2-dimensional case. The dotted lines denote the maxima of the respective curves. Increasing n or ρ_τ shifts the the respective maximum to the right.

4.3.1. Estimation

Specifying $\hat{m}_{\mathsf{p}}$ requires us to choose an appropriate estimator for θ, which we denote by $\hat{\theta}$. Let $\hat{\theta} \sim N(\mu_{\hat{\theta}}, \sigma^2_{\hat{\theta}})$.[7] The expected value and variance of $\hat{m}_{\mathsf{p}}$ are then given by:

$$\begin{aligned} \mathrm{E}(\hat{m}_{\mathsf{p}}) &= n^{\frac{2}{d+4}} \, \mathrm{E}\left(\hat{\theta} e^{\hat{\theta}\frac{1}{n}}\right) \\ &= n^{\frac{2}{d+4}} \left(\mu_{\hat{\theta}} + \frac{\sigma^2_{\hat{\theta}}}{n}\right) e^{\frac{1}{n}\left(\mu_{\hat{\theta}} + \frac{\sigma^2_{\hat{\theta}}}{2n}\right)} \end{aligned} \tag{4.25}$$

and

$$\begin{aligned} \mathrm{Var}(\hat{m}_{\mathsf{p}}) &= n^{\frac{4}{(d+4)^2}} \, \mathrm{Var}\left(\hat{\theta} e^{\hat{\theta}\frac{1}{n}}\right) \\ &= n^{\frac{4}{(d+4)^2}} \frac{1}{n^2} \left[\mu^2_{\hat{\theta}} n^2 \left(e^{\frac{\sigma^2_{\hat{\theta}}}{n^2}} - 1\right) + 4\mu_{\hat{\theta}} n \left(\sigma^2_{\hat{\theta}} e^{\frac{\sigma^2_{\hat{\theta}}}{n^2}} - 2\sigma^2_{\hat{\theta}}\right)\right. \\ &\quad \left. + n^2 \sigma^2_{\hat{\theta}} e^{\frac{\sigma^2_{\hat{\theta}}}{n^2}} \left(n^2 + 4\sigma^2_{\hat{\theta}}\right) - \sigma^4_{\hat{\theta}}\right] e^{\frac{1}{n}\left(2\mu_{\hat{\theta}} + \frac{\sigma^2_{\hat{\theta}}}{n}\right)}. \end{aligned} \tag{4.26}$$

From (4.25) we see that the fractions $\frac{\sigma^2_{\hat{\theta}}}{n}, \frac{\sigma^2_{\hat{\theta}}}{2n}$ vanish for $n \to \infty$. Thus $\hat{m}_{\mathsf{p}}$ is asymptotically unbiased if $\mu_{\hat{\theta}} = \theta$. In the following we show that $\hat{m}_{\mathsf{p}}$ is unbiased for $\hat{\theta} = \hat{\rho}_\tau$. Kendall's tau is defined by:

$$\rho_\tau = \mathrm{E}\left[\mathrm{sgn}\left(\left(X_1 - \tilde{X}_1\right)\left(X_2 - \tilde{X}_2\right)\right)\right], \tag{4.27}$$

where $(X_1, X_2)^\top$ is a vector of two continuous random variables and $(\tilde{X}_1, \tilde{X}_2)^\top$ is an independent copy of $(X_1, X_2)^\top$.
Let $(X_i, X_j)_n$ denote n pairwise *i.i.d.* observations. The estimator for ρ_τ is given by:

$$\hat{\rho}_\tau(X_i, X_j) = \binom{n}{2}^{-1} \sum_{i \le s < t \le n} \mathrm{sgn}\left(\left(X_{s,i} - X_{t,i}\right)\left(X_{s,j} - X_{t,j}\right)\right), \tag{4.28}$$

where the summation is taken over the $\binom{n}{2}$ pairs of n. We note that, since the sample is *i.i.d.* we have, without loss of generality:

$$\begin{aligned} \mathrm{E}\left[\mathrm{sgn}\left(\left(X_{s,i} - X_{t,i}\right)\left(X_{s,j} - X_{t,j}\right)\right)\right] &= \mathrm{E}\left[\mathrm{sgn}\left(\left(X_{s,1} - X_{t,1}\right)\left(X_{s,2} - X_{t,2}\right)\right)\right] \\ &= \mathrm{E}\left[\mathrm{sgn}\left(\left(X_1 - \tilde{X}_1\right)\left(X_2 - \tilde{X}_2\right)\right)\right]. \end{aligned}$$

[7] This assumption is non critical for appropriately designed estimators, this follows from empirical process theory. See e.g. Van der Vaart, 2000.

Therefore:

$$\begin{aligned}\mathrm{E}(\hat{\rho}_\tau) &= \binom{n}{2}^{-1} \sum_{1\le s<t\le n} \mathrm{E}\left[\operatorname{sgn}\left((X_{s,i}-X_{t,i})\,(X_{s,j}-X_{t,j})\right)\right] \\ &= \binom{n}{2}\binom{n}{2}^{-1} \mathrm{E}\left[\operatorname{sgn}\left(\left(X_1-\tilde{X}_1\right)\left(X_2-\tilde{X}_2\right)\right)\right] \\ &= \rho_\tau. \end{aligned} \tag{4.29}$$

Furthermore, we state without proof:[8]

$$\begin{aligned}\mathrm{Var}(\hat{\rho}_\tau) &= \binom{n}{2}^{-2} \mathrm{Var}\left[\sum_{1\le s<t\le n} \operatorname{sgn}\left(\left(X_1-\tilde{X}_1\right)\left(X_2-\tilde{X}_2\right)\right)\right] \\ &= \binom{n}{2}^{-1} \sum_{c=1}^{2} \binom{2}{c}\binom{n-2}{2-c}\sigma_c^2. \end{aligned} \tag{4.30}$$

with

$$\sigma_c^2 = \mathrm{Var}\left[\operatorname{sgn}\left((X_{s,1}-X_{t,1})\,(X_{s,c}-X_{t,c})\right)\right].$$

It is clear that since $\hat{\theta}$ is the only random variable in $\hat{m}_{\mathsf{p}}$, the estimation of this criterion reduces to the estimation of the dependence measure. Thus, in the case of choosing Kendall's tau we can easily derive the consistency of $\hat{m}_{\mathsf{p}}$ from the consistency of $\hat{\rho}_\tau$. The latter follows from the Equations (4.29) and (4.28) by the law of large numbers.

4.3.2. Specifying the concordance measure

Before we proceed with the analysis we have to consider additional consequences of parameterizing our criterion with a concordance measure ρ_ξ. Concordance measures are measures of association, that quantify the degree at which a pair of random variables (X, Y) tend to move together, i.e. whether we can state that "large" values of one variable tend to occur with "large" values of the other and "small" values tend to occur with "small" values of the other. This information is quantified differently by concordance measures such as Kendall's tau or Spearman's rho. We have seen from (4.27) that the former measures the difference in probability of the occurrence of concordant pairs to discordant pairs of a bivariate sample. The interpretation here is clearly, that if the former is more probable than the latter ρ_τ is positive,

[8] see e.g. Lehmann, 1999, pp. 368-369. Kendall's tau is a U-statistic of degree 2. This expression is derived under relatively mild conditions, namely that the variance of any subset of the random vector $\mathbf{X}$ exists and is finite.

otherwise negative. We have seen in Chapter 2 that Spearman's rho however can be expressed as:

$$\rho_S = 12 \int \int [C(u,v) - uv]\, dudv,$$

thus measuring a scaled distance between the distribution of (u, v) and the independence copula $\prod$. These are two fundamentally different approaches at quantifying the association between X and Y. Spearman's rho is in this sense identical to Pearson's correlation coefficient, with the exception that it is based on the rank-transformed variables $(F(X), F(Y)) = (U, V)$. This can be seen from the following representation:[9]

$$\rho_S = \frac{\mathrm{E}(U,V) - \mathrm{E}(U)\,\mathrm{E}(V)}{\sqrt{\mathrm{Var}(U)}\sqrt{\mathrm{Var}(V)}}$$

Consequently Kendall's tau should be interpreted as the difference in *probability* of the occurrence of concordant to discordant pairs, whereas Spearman's rho quantifies the distance in probability mass to what could be expected under independence.

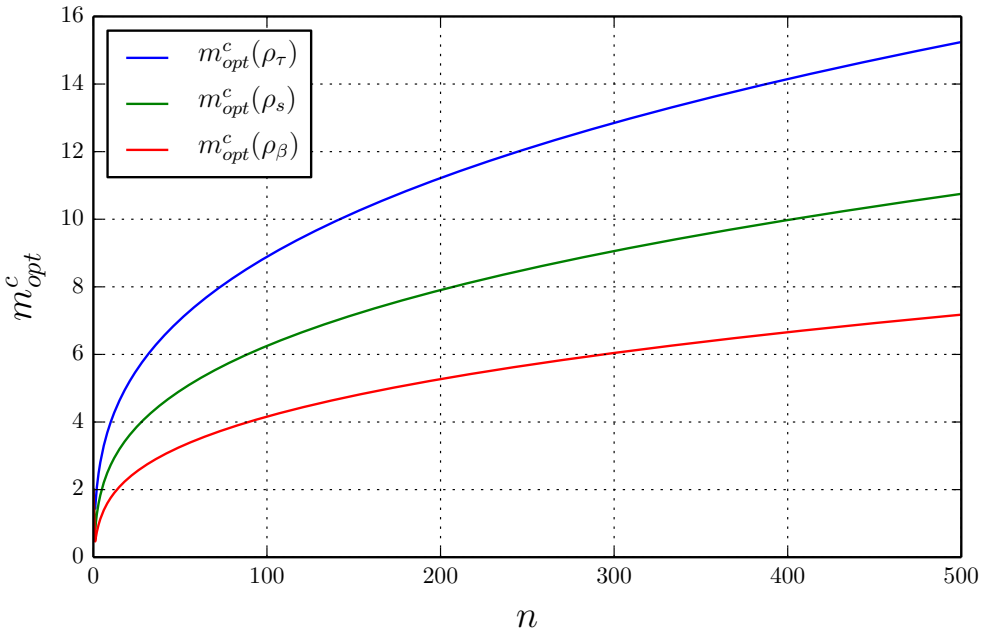

Figure 4.4.: $\hat{m}_{\mathsf{p}}$ parameterized with Kendall's tau, Spearman's rho and Blomqvist's beta for a bivariate gauss copula with $\boldsymbol{\mu} = (0,0)^\top$ and $\rho = 0.5$.

Another measure that satisfies the assumptions we made in Equations (4.13) -

[9] see e.g. Nelsen, 2007.

(4.15) is Blomqvist's β:

$$\beta = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1,$$

which is constructed similarly to Kendall's tau, as can be seen from:[10]

$$\beta = P\left[(X - \tilde{x})(Y - \tilde{y}) > 0\right] - P\left[(X - \tilde{x})(Y - \tilde{y}) < 0\right]$$

where $\tilde{x}$ and $\tilde{y}$ denote the medians of X and Y. In Figure 4.4 we've plotted these three measures for a bivariate Gauss copula with $\mu = (0, 0)^\top$ and $\rho = 0.5$. We can observe that the consequences of selecting e.g. Kendall's tau over Blomqvist's β can be significant. Since $\hat{m}_{\mathsf{p}}$ is specified with the absolute value of Θ the orderings do not depend on the sign of the dependence measure. There seems to be no objective reason to discriminate between either of the presented concordance measures, as each lend themselves to meaningful interpretations, but since Kendall's tau and Spearman's rho are widely used in practice we propose to parameterize $\hat{m}_{\mathsf{p}}$ with the empirical average, that is constructed from both measures:

$$\rho_{\tau,s} = \frac{1}{2}(\rho_\tau + \rho_s).$$

4.3.3. Simulation study: 2-dimensional case

We conducted a monte carlo simulation study to identify an optimal grid size m_{opt}, which we compared to the estimators $\hat{m}_{\mathsf{p}}$ and $\hat{m}^{lscv}_{opt}$. We simulated n bivariate samples $\{U_{i.1}, U_{i,2}\}^n_i$, where $n \in \{50, 60, \ldots, 500\}$ for the Gumbel- (C^{Gu}_θ), Clayton- (C^{Cl}_θ), Frank- (C^{Fr}_θ), Gauss- (C^{Ga}_ρ), t- ($C^{t}_{\nu,\rho}$) and Joe-(C^{Jo}_θ). We computed the grid size m^n_{opt} that minimizes the estimated mean integrated squared error for the respective sample $\widehat{MISE}_n$ in brute force optimizations for each copula we specified and for individual parameter ranges per copula. The per sample estimator $\widehat{MISE}_n$ is defined as[11]:

$$\begin{aligned}
\widehat{MISE}_n &= \widehat{IBias}^2 + \widehat{I\,\mathrm{Var}} \\
&\approx \frac{1}{I}\sum_{k=1}^{I}\left[\left(\frac{1}{N}\sum_{j=1}^{N}\hat{c}_j(s_k) - c(s_k)\right)^2\right] \\
&\quad + \frac{1}{I}\sum_{k=1}^{I}\left[\frac{1}{N}\sum_{j=1}^{N}\left(\hat{c}_j(s_k) - \frac{1}{N}\sum_j \hat{c}(s_k)\right)^2\right],
\end{aligned}$$

[10] see e.g. Nelsen, 2007.

[11] see Bouezmarni et al., 2013.

where $I = 50, N = 1000$ and $s_k = (u_{1,k}, u_{2,k})$ are randomly chosen points from the independence copula $u_1 u_2$. $\hat{c}_j(s)$ for $j \in \{1, \dots N\}$ corresponds to the jth Bernstein copula density estimation.

The relationship between ρ_τ and $\hat{m}_{\mathsf{p}}$ can be seen on all datasets. At levels of Kendall's tau near 0, m_{opt}^{MSE} is estimated at its lowest possible value across almost all datasets, whereas at higher levels $\hat{m}_{\mathsf{p}}$ attains correspondingly higher values. Table 4.1 summarizes descriptive statistics with respect to the estimates of the MSE-optimal grid sizes $\hat{m}_{opt}^{\text{MSE}}$. Since $\hat{m}_{opt}^{lscv}$ hasn't always attained its minimum within the specified domain, we chose to use its value at the boundary of the interval as its minimum value in these cases. This adds a level of arbitrariness to the respective statistics, but we consider this procedure reasonable, with regard to the fact, that the estimate of the MSE-optimal grid size attains its minimum within this range. We thus underestimate the respective statistics in cases where the bounds are attained. There is strong evidence, that m_{opt}^{MSE} is increasing in the absolute value of Kendall's tau $|\rho_\tau|$, which supports our modeling decisions. The latter can be seen in a more compact way by the last sub-table. Here we have split the data into a center, middle and outer region CR, MR and $, OR$, that are comprised of the intervals $[-0.1, 0.1], \{[-0.5, -0.3], [0.3, 0.5]\}$ and $\{[-0.9, -0.7], [0.7, 0.9]\}$, respectively. As we move from the center to the outer region $\hat{m}_{opt}^{\text{MSE}}$ increases everywhere in this reduced dataset. The t-copula, which we parameterized with 4 degrees of freedom seems to break this structure, as the values in the center region are significantly higher than compared to the other copulas. But remembering that a t-copula with $\nu = 4$ allocates significant amount of probability mass to all 4 corners of its support, while allocating the rest evenly to the rest of the domain, we can explain these values. For lower sized samples the probability mass in the corners is less pronounced, so the evenly spaced sector of the supports dominates the dependence structure, i.e. for small samples with this parameterization we are close to the independence copula. Whereas for larger samples, this effect is reversed and realizations in the corners are more frequent, which leads to high values of m_{opt}^{MSE}.

The simulation results clearly indicate, that the cross validation estimate overestimates the optimal polynomial order systematically in the sample range. To quantify the goodness of fit of the respective models we chose to measure the deviation of the respective estimates from $\theta_{opt}^{\text{MSE}}$ with the root mean squared error $RMSE$:

$$RMSE = \sqrt{\text{MSE}} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - f_i)\,,}{n}}$$

Even though this measure seems to have deficiencies with regard to its reliability, we still report the RMSE as it is widely known and used in the scientific community.[12]

[12]For a discussion of various error measures in econometrics see e.g. Armstrong and Collopy, 1992.

Table 4.1.: Descriptive statistics for $\hat{m}_{opt}^{\text{MSE}}$ - 2-Dimensional data.

	C_{ρ}^{Ga}		$C_{\nu,\rho}^{t}$		C_{θ}^{Fr}	
ρ_τ	$mean$	CI	$mean$	CI	$mean$	CI
-0.9	19.80	(16.21, 23.3)	21.57	(16.56, 26.15)	13.04	(8.6, 18.26)
-0.7	8.63	(7.34, 10.19)	19.76	(14.82, 24.6)	9.43	(5.86, 14.19)
-0.5	5.57	(5.19, 5.91)	21.76	(16.63, 26.17)	5.67	(3.45, 9.34)
-0.3	4.20	(3.89, 4.43)	14.22	(9.82, 19.13)	2.74	(2.0, 4.21)
-0.1	2.61	(2.43, 2.71)	12.30	(7.86, 16.73)	2.00	(2.0, 2.0)
0.1	2.00	(2.0, 2.0)	3.48	(2.0, 5.69)	2.00	(2.0, 2.0)
0.3	4.22	(3.86, 4.47)	4.63	(3.93, 5.91)	2.96	(2.82, 2.97)
0.5	5.43	(5.04, 5.82)	6.52	(5.8, 7.41)	2.00	(2.0, 2.0)
0.7	8.09	(6.97, 9.23)	10.46	(8.67, 12.86)	8.59	(7.19, 10.28)
0.9	22.96	(18.76, 26.6)	18.57	(15.3, 22.08)	2.02	(2.0, 2.06)
ρ_τ		C_{θ}^{Cl}		C_{θ}^{Gu}		C_{θ}^{Jo}
	$mean$	CI	$mean$	CI	$mean$	CI
0.1	2.02	(2.0, 2.06)	2.52	(2.34, 2.65)	2.07	(2.0, 2.19)
0.3	2.72	(2.52, 2.86)	3.96	(3.6, 4.23)	2.72	(2.5, 2.93)
0.5	5.65	(4.71, 8.28)	4.85	(4.54, 5.28)	5.70	(4.82, 6.95)
0.7	10.80	(9.02, 12.84)	9.24	(8.26, 10.15)	11.59	(9.45, 14.17)
0.9	19.85	(16.06, 23.8)	19.33	(16.54, 22.78)	21.78	(17.67, 25.67)
	C_{ρ}^{Ga}	$C_{\nu,\rho}^{t}$	C_{θ}^{Fr}	C_{θ}^{Cl}	C_{θ}^{Gu}	C_{θ}^{Jo}
			$\sum_R mean$			
CR	4.61	15.78	4.00	2.02	2.52	2.07
MR	19.41	47.13	13.37	8.37	8.80	8.41
OR	59.48	70.35	33.09	30.65	28.57	33.37

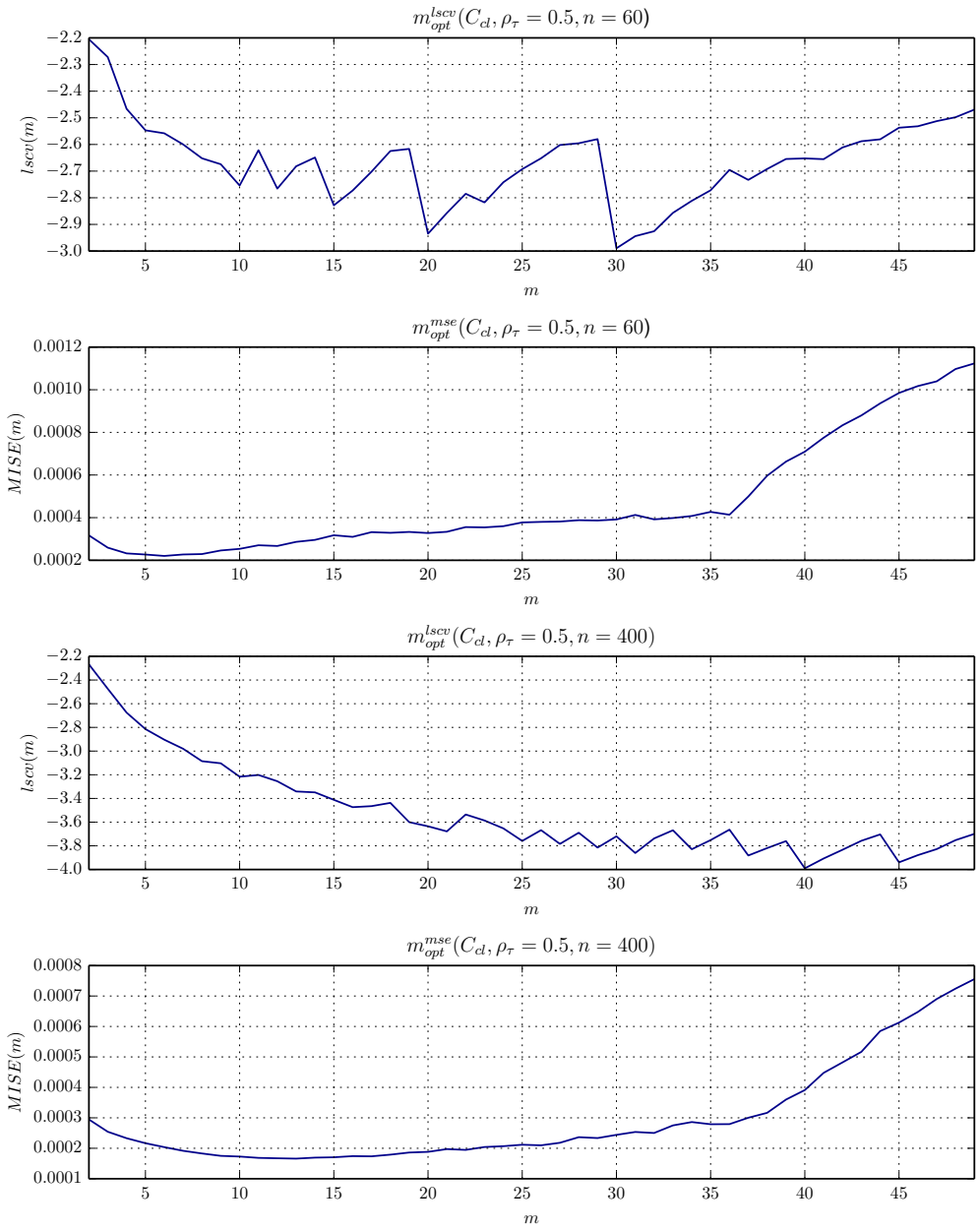

Figure 4.5.: Plots of lscv(m) against m. The figures show cases of existence and non-existence of a minimum within the specified range of grid sizes for m_{opt}^{lscv}. In contrast, m_{opt}^{MSE} always attains its minimum in the same range.

In Appendix A.2 we also report the Mean Absolute Scaled Error ($MASE$) of the

simulation data, where the scaled error is defined as

$$q_t = \frac{e_t}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_t - Y_{t-1}|}.$$

Taking the absolute value of the scaled error and averaging across the dataset then gives:

$$MASE = \frac{1}{n-1}\sum |q_t|.$$

This measure, introduced by Hyndman and Koehler (2006), thus quantifies the measurement error by evaluating the forecast error in relation to the mean prediction error of a random walk. The results differ marginally.
Table 4.2 summarizes the $RMSE$ between $\hat{m}_{opt}^{\text{MSE}}$ and each estimator. $\hat{m}_{\mathsf{p}}$ captures the relationship, that the optimal polynomial order depends on the strength of concordance within the random vector. In contrast, $\hat{m}_{opt}^{lscv}$ only uses information that is inherently a property of the Bernstein copula which is not sufficient for medium sized samples.

In Table 4.3 we report the median values of the mean squared errors at the respective optimal grid sizes across the sample size interval $[50, 60, \ldots, 500]$ to approximate the precision of a Bernstein copula estimate that is parameterized by $\hat{m}_{\mathsf{p}}$.

4.3.4. Detecting Spearman's rho

In addition to the aforementioned simulation we conducted the following study to analyze the extent to which the Bernstein density estimator, parameterized with $\hat{m}_{\mathsf{p}}$ and m_{opt}^{lscv} respectively, captures central properties related to the estimand. In this simulation we generate samples from the Gauss copula for the parameter range $\rho \in \{-0.9, -0.7, \ldots, 0.9\}$, and calculate the corresponding Bernstein copulas. Since rank correlations are copula parameters we can compute Spearman's rho for each Bernstein copula and compare it with the true value that generated the respective sample. For the Gauss copula the true value can be computed by the well known formula[13]:

$$\rho_s(X_1, X_2) = \frac{6}{\pi}\arcsin\frac{1}{2}\rho \tag{4.31}$$

Spearman's rho with respect to the Bernstein copula is defined as[14]:

$$\rho_s(X_1, X_2) = 12\sum_{v_1=0}^{m}\sum_{v_2=0}^{m}\gamma\left(\frac{v_1}{m}, \frac{v_2}{m}\right)\prod_{j=1,2}\binom{m}{v_j}B(v_j + 1, m + 1 - v_j), \tag{4.32}$$

[13] see e.g. McNeil et al., 2010.
[14] see Sancetta and Satchell, 2004.

Table 4.2.: $RMSE$ across the sample range $[50, 60, \ldots, 500]$ per parameter for 2-dimensional simulated data. The random samples were generated with the parameters that correspond to the respective Kendall's tau values. For C_θ^{Cl} and C_θ^{Gu} negative values of ρ_{tau} either lead to non-strict copulas or, as in the latter case, are not defined.

$RMSE$							
			$\hat{m}_{\mathsf{p}}$				
ρ_τ	C_ρ^{Ga}	$C_{\nu,\rho}^{t}$	C_θ^{Fr}	C_θ^{Cl}	C_θ^{Gu}	C_θ^{Jo}	$\frac{1}{n}\sum$
-0.9	12.67	17.26	17.24	n.a.	n.a.	n.a.	15.72
-0.7	5.31	18.58	14.87	n.a.	n.a.	n.a.	12.92
-0.5	3.07	20.75	11.74	n.a.	n.a.	n.a.	11.85
-0.3	0.99	18.24	5.47	n.a.	n.a.	n.a.	8.23
-0.1	0.95	18.77	0.45	n.a.	n.a.	n.a.	6.72
0.1	0.45	7.12	0.45	0.51	0.88	0.64	1.67
0.3	0.99	3.04	2.33	2.64	1.26	2.66	2.15
0.5	3.21	3.18	6.73	5.47	3.89	4.37	4.48
0.7	4.88	6.38	5.85	6.43	3.77	8.20	5.92
0.9	15.07	11.55	13.68	14.37	11.52	14.41	13.43
$\frac{1}{n}\sum$	4.76	12.49	7.88	5.88	4.26	6.05	8.31
			$\hat{m}_{opt}^{lscv}$				
ρ_τ	C_ρ^{Ga}	$C_{\nu,\rho}^{t}$	C_θ^{Fr}	C_θ^{Cl}	C_θ^{Gu}	C_θ^{Jo}	$\frac{1}{n}\sum$
-0.9	29.12	29.12	19.00	n.a.	n.a.	n.a.	25.75
-0.7	32.21	32.41	23.91	n.a.	n.a.	n.a.	29.51
-0.5	33.75	35.63	27.36	n.a.	n.a.	n.a.	32.24
-0.3	35.21	38.12	32.18	n.a.	n.a.	n.a.	35.17
-0.1	30.22	40.92	43.29	n.a.	n.a.	n.a.	38.14
0.1	31.20	41.01	29.67	31.09	24.90	32.22	31.68
0.3	32.87	38.16	32.27	34.68	31.62	36.44	34.34
0.5	33.70	35.74	17.90	35.06	34.72	35.51	32.10
0.7	32.11	32.54	31.80	32.40	32.28	32.41	32.26
0.9	29.12	29.12	18.14	29.1	29.12	29.12	27.29
$\frac{1}{n}\sum$	31.95	35.28	27.55	32.47	30.53	33.14	31.85

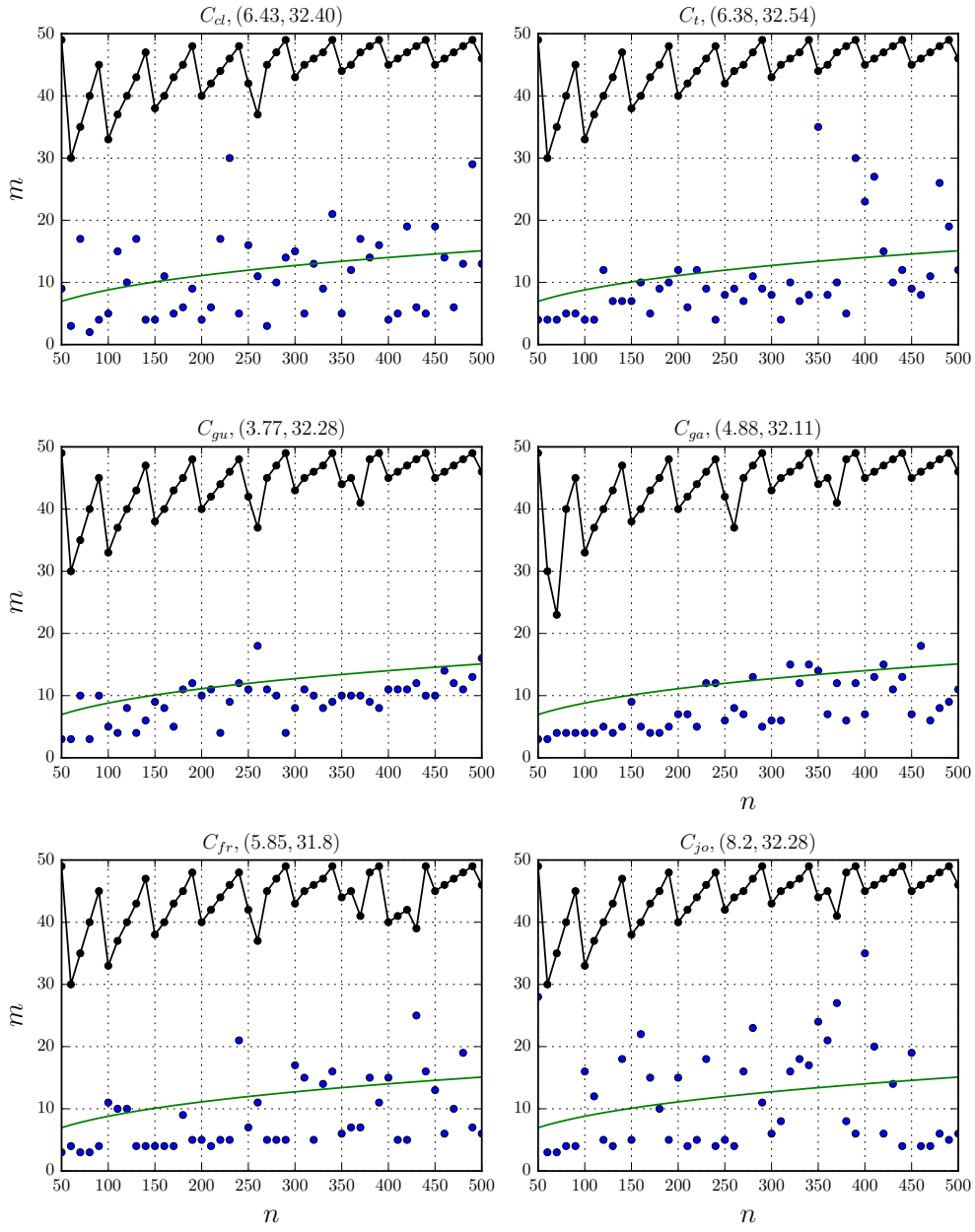

Figure 4.6.: Optimal grid sizes in 2 dimensions plotted against $\hat{m}_{\mathsf{p}}$ and $\hat{m}_{opt}^{lscv}$, $\theta = \rho_\tau = 0.7$. The blue dots show the MSE-optimal grid sizes, the green and black lines represent $\hat{m}_{\mathsf{p}}$ and $\hat{m}_{opt}^{lscv}$, respectively. The $RMSE$ with respect to $\hat{m}_{opt}^{\text{MSE}}$ and $\hat{m}_{\mathsf{p}}$ is reported in brackets.

where

$$\gamma\left(\frac{v_1}{m}, \frac{v_2}{m}\right) = \hat{C}\left(\frac{v_1}{m}, \frac{v_2}{m}\right) - \frac{v_1}{m}\frac{v_2}{m}. \tag{4.33}$$

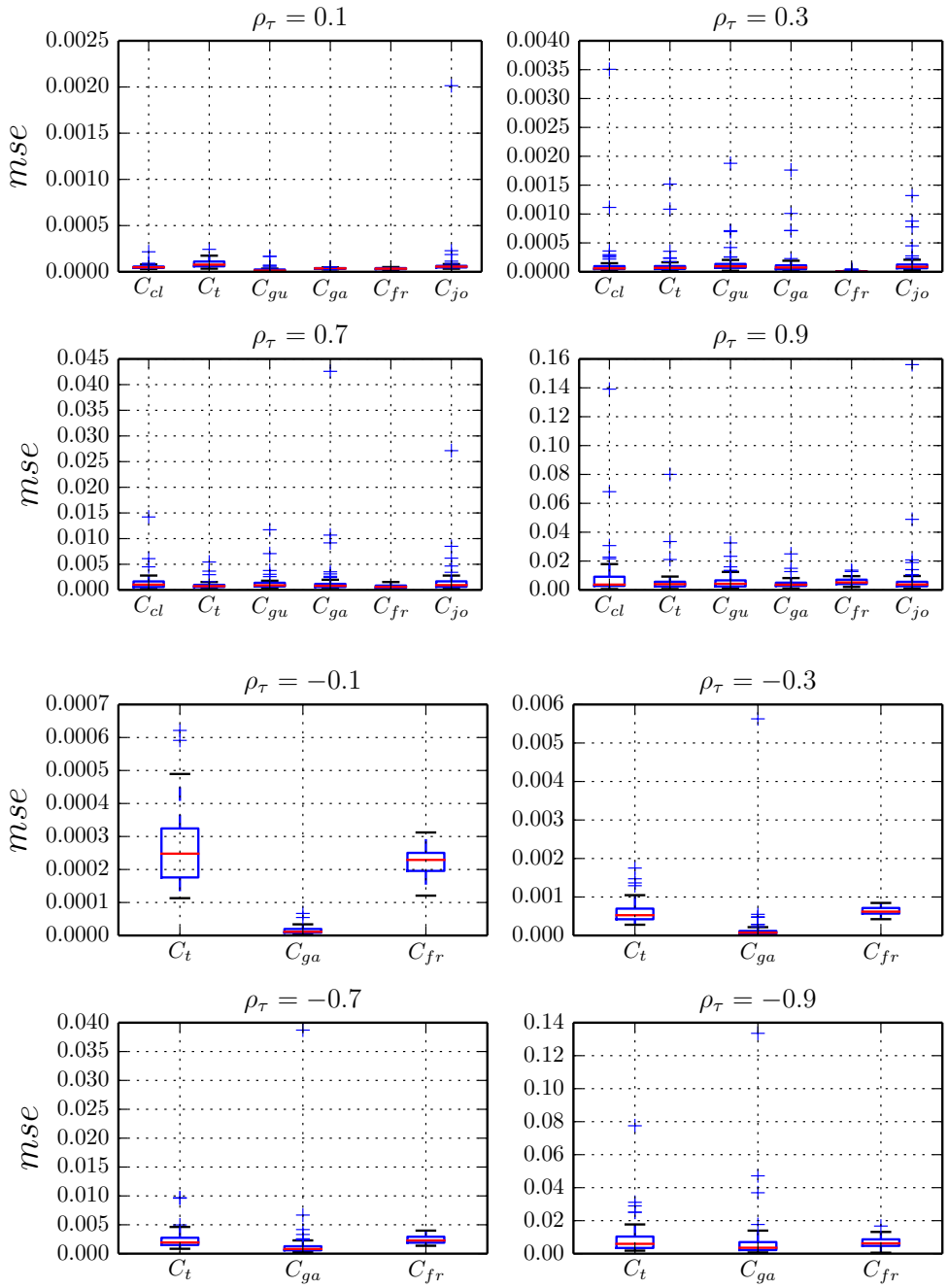

Figure 4.7.: Box plots of mean squared errors for selected levels of Kendall's tau in the 2-dimensional case. The mean squared errors correspond to the respective optimal grid sizes in the sample size interval $[50, 60, \ldots, 500]$.

Table 4.3.: Median mean squared errors for 2-Dimensional data . For each value of ρ_τ the median is obtained with respect to the mean squared errors of each sample size in the interval $[50, 60 \ldots .500]$.

ρ_τ	C_{ga}	C_t	C_{fr}	C_{cl}	C_{gu}	C_{jo}
-0.9	0.003575	0.005945	0.006246	n.a.	n.a.	n.a.
-0.7	0.000802	0.001930	0.002289	n.a.	n.a.	n.a.
-0.5	0.000182	0.001156	0.001142	n.a.	n.a.	n.a.
-0.3	0.000073	0.000527	0.000620	n.a.	n.a.	n.a.
-0.1	0.000011	0.000248	0.000229	n.a.	n.a.	n.a.
0.1	0.000037	0.000077	0.000034	0.000049	0.000013	0.000019
0.3	0.000075	0.000068	0.000005	0.000059	0.000094	0.000005
0.5	0.000222	0.000202	0.000583	0.000217	0.000247	0.000006
0.7	0.000805	0.000738	0.000599	0.001041	0.000907	0.000079
0.9	0.003607	0.004012	0.005222	0.003706	0.004271	0.001490

4.3.5. Simulation study: n-dimensional case, n > 2

In higher dimensions we have to specify a suitable mechanism to reduce the pairwise rank-correlations of $\mathbf{X}$ to a one-dimensional criterion that captures the required dependency information in a way that is equivalent to θ.[15][16]A natural specification would set θ to the weighted average of the respective pairwise realizations:

$$\theta = \binom{d}{2}^{-1} \sum_{\binom{d}{2}} \theta_{i,j} \mathbb{1}_{\{i \neq j\}}, i, j \in \{1, \ldots d\}. \tag{4.34}$$

Multivariate extensions of Spearman's rho of this form are reviewed e.g. in Schmid and Schmidt (2007). With respect to this dependence measure the above estimator is referred to as population version of the average pair-wise Spearman's rho which the authors define by:

$$\rho_3 = 3 \left[4 \sum_{k<l} \binom{d}{2}^{-1} \int_{[0,1]^2} C_{kl}(u, v) du dv - 1 \right] \tag{4.35}$$

A disadvantage of this specification is that in the presence of rank correlations of opposite signs these values would cancel each other out, which is counter-intuitive in situations when the respective pairwise rank correlations show strong dependency.

[15] see e.g. Joe, 1990, for a discussion on multivariate concordance measures.
[16] see e.g. Jaworski et al., 2010, pp. 209 for a more recent discussion on this topic.

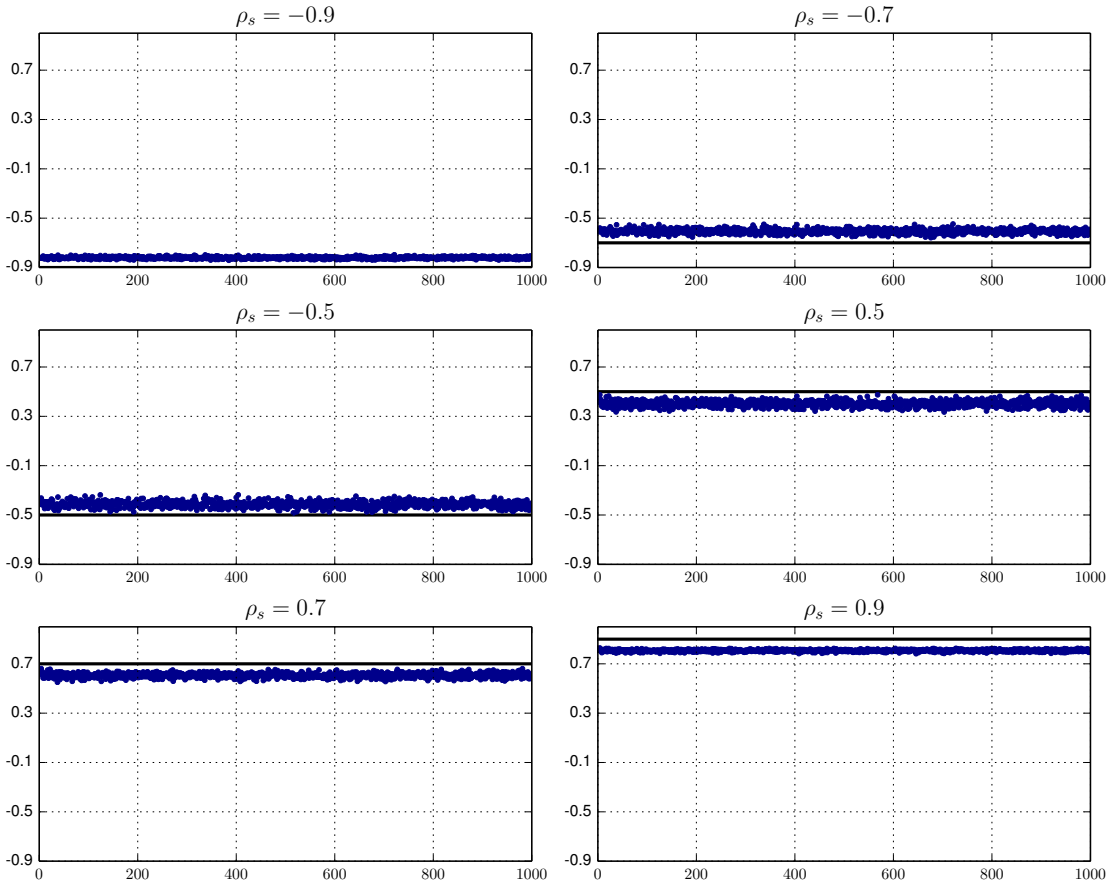

Figure 4.8.: Estimation of Spearman's rho at sample size 800.

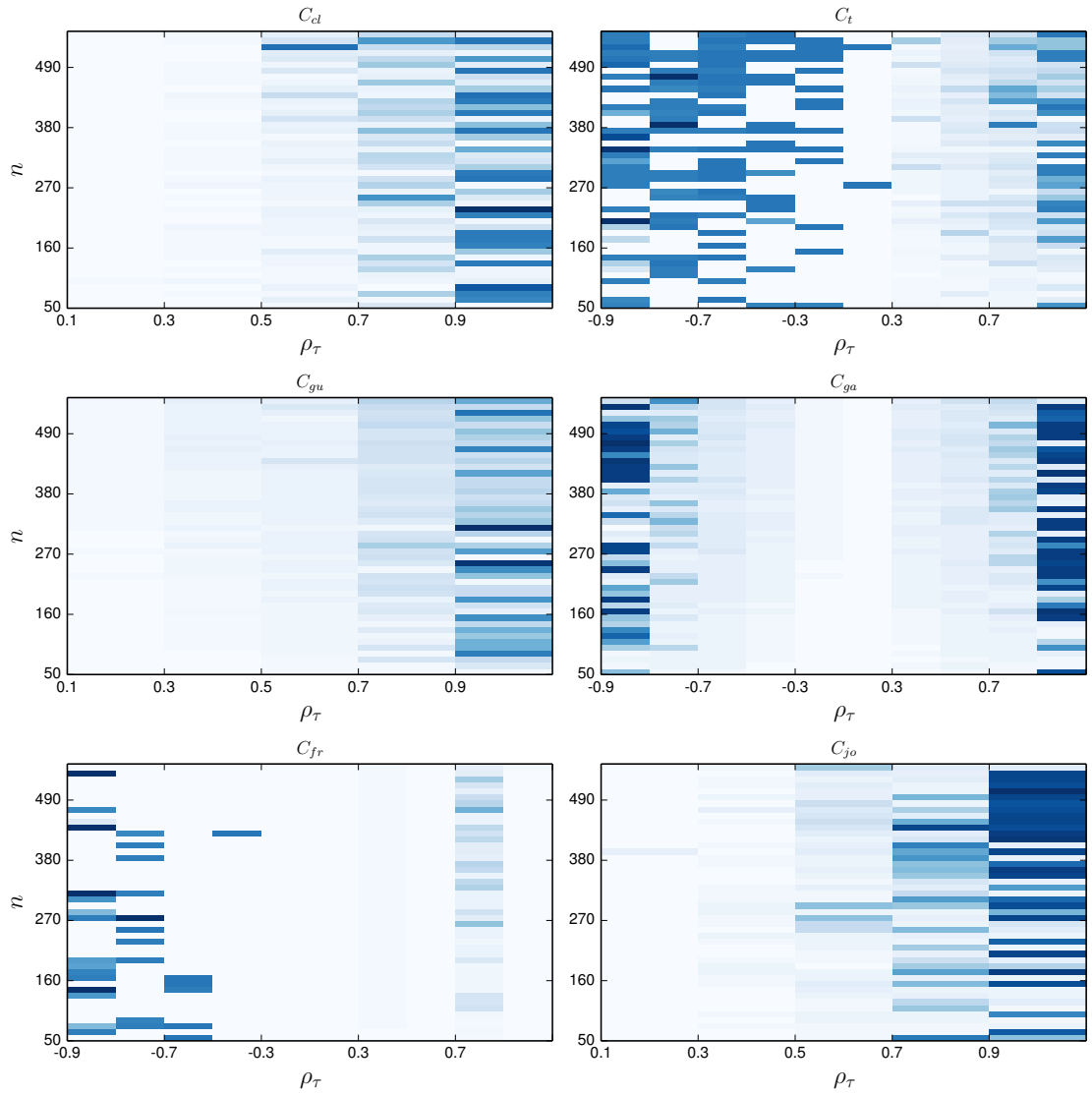

Figure 4.9.: Relationship between n, ρ_τ and m_{opt}^{MSE}. Dark colors indicate a high optimal grid size.

The authors further present the two following extensions:

$$\rho_1 = h(d) \left[2^d \int_{[0,1]^d} C(u)du - 1 \right] \tag{4.36}$$

$$\rho_2 = h(d) \left[2^d \int_{[0,1]^d} \prod(u)dC(u) - 1 \right]. \tag{4.37}$$

where

$$h(d) = \frac{d+1}{2^d - (d+1)}.$$

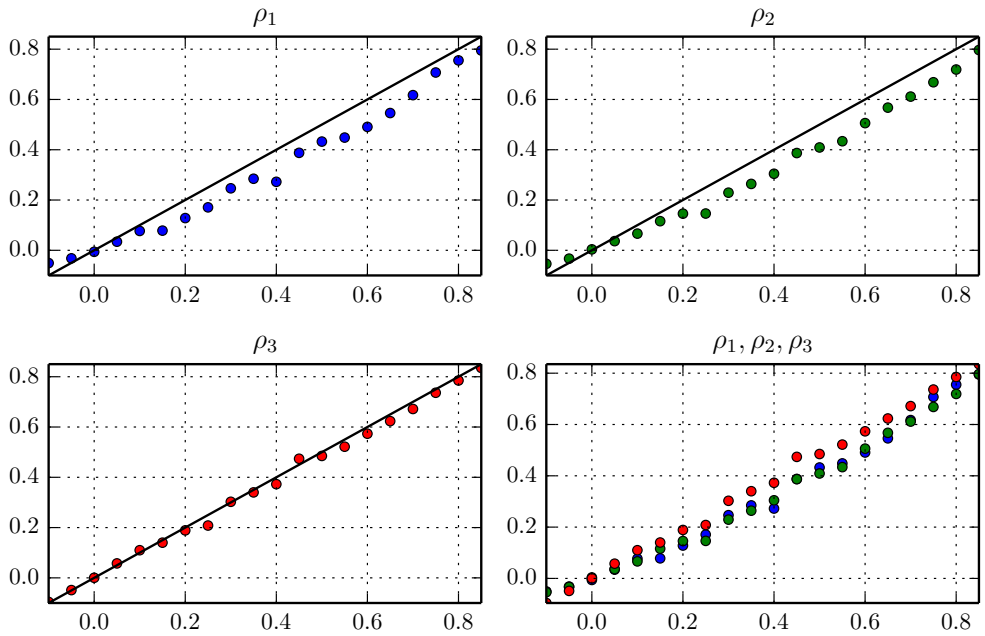

Figure 4.10.: ρ_1, ρ_2 and ρ_3 for a 6-Dimensional random vector $\mathbf{U} = (U_1, \ldots, U_6)^\top$ taken from a Gauss-Copula with $\rho(U_i, U_j) = \rho$, $\forall i \neq j$. The solid line represents the specification (4.34).

Another possible strategy is to select the maximum of the absolute value of the respective dependence measures:

$$\theta = \max_{\binom{d}{2}} |\theta_{i,j}|, i, j \in \{1 \ldots d\} \wedge i \neq j. \tag{4.38}$$

An argument for (4.38) can be made from the perspective of dimension reduction techniques like principal component analysis or factor models, where high-dimensional problems are reduced to subsets of lower dimension. The essential argument here is that in the case, that the randomness in a random vector is mainly driven by a smaller subset of the same vector, one can ignore the irrelevant components and focus on the low dimensional subset for further analysis. In the current context one could argue that given a random vector X and its rank correlation matrix $\rho_\xi(X)$ it might be feasible to focus on the maximum of all rank correlations as it has the most influence on the joint dependency structure.

An intuition for the validity of this specification may be gained by Figure 4.11.

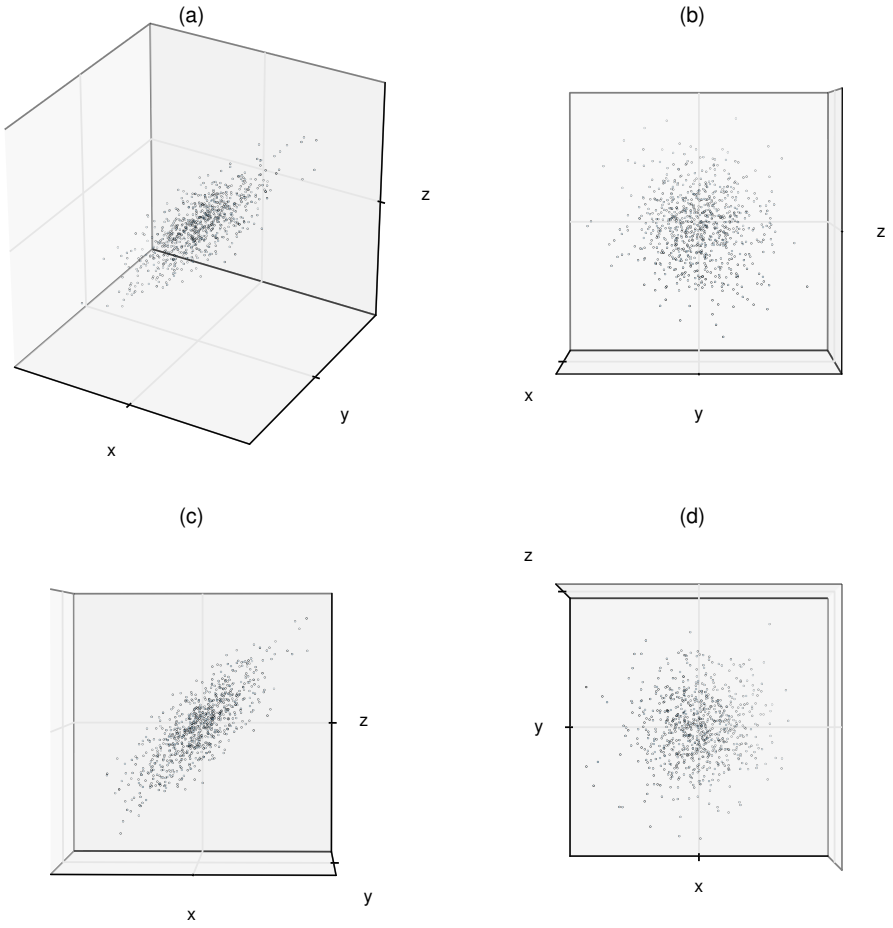

Figure 4.11.: Perspective scatter plots of $(x, y, z) \sim N(\mu, \mathbf{\Sigma})$, where $\rho(x, y) = 0.1, \rho(x, z) = 0.8, \rho(y, z) = 0.1$. **(a)** shows the general elliptical dependence structure, **(b)**, **(c)** and **(d)** focus on different angles of the same plot to emphasize the correlation of the pairs $(y, z), (x, z)$, and (x, y), respectively.

The graphic shows a random vector $(x, y, z) \sim N(\mu, \mathbf{\Sigma})$ with correlation structure $\rho(x, y) = 0.1, \rho(x, z) = 0.8, \rho(y, z) = 0.1$. As we rotate the graphic to focus on the correlation structure of the 3 pairs $(y, z), (x, z), (x, y)$, we can clearly observe that $\rho(x, z)$ in a sense dominates the dependence structure, as it induces the elliptic shape into a scatter plot, that would otherwise be spherically shaped.

We generated 3-dimensional data according to the same specifications and for the same copulas we introduced in the previous section. To keep the computational time

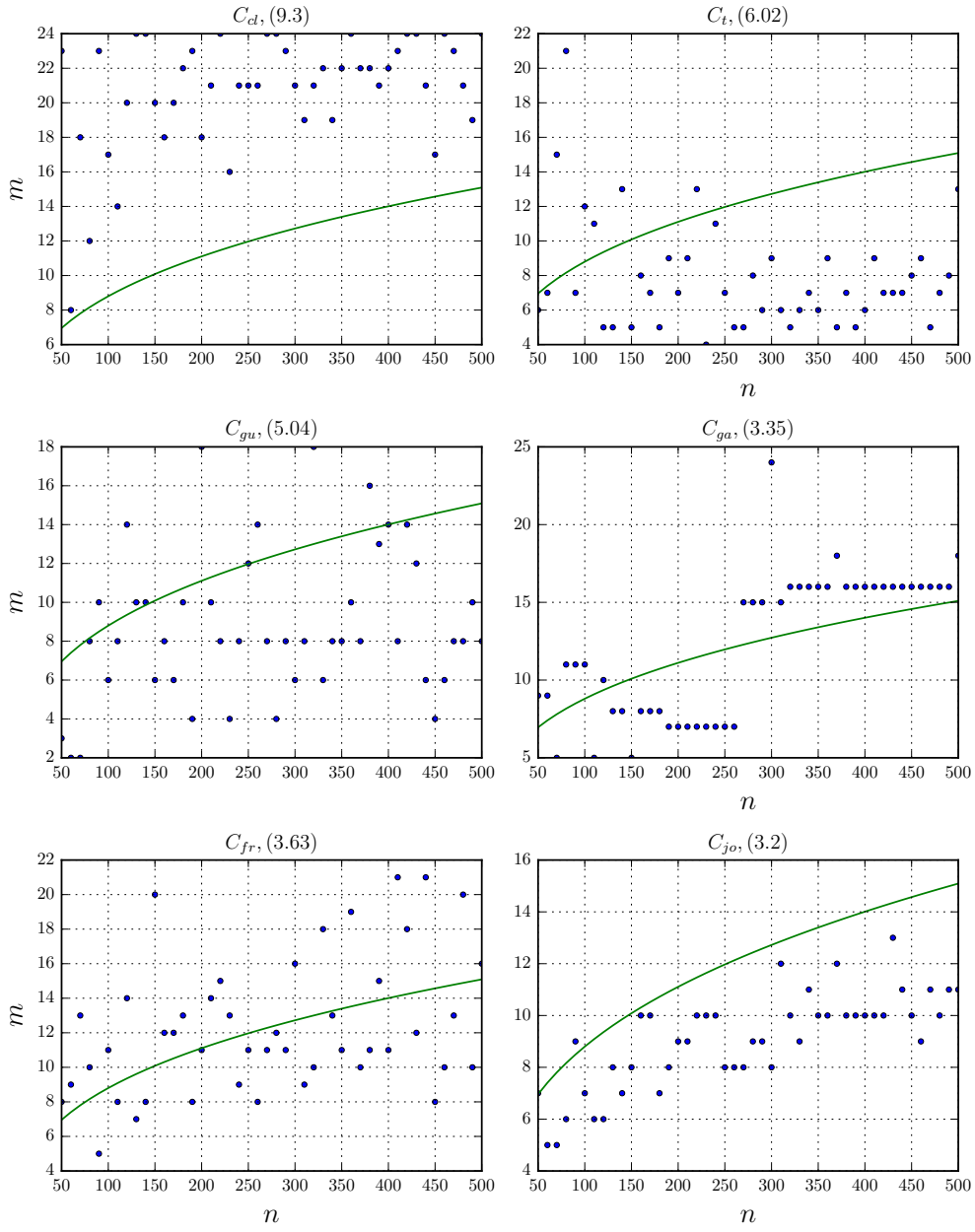

Figure 4.12.: Optimal grid sizes in 3 dimensions plotted against $\hat{m}_{\mathsf{p}}$, $\theta = \rho_\tau = 0.7$. The blue dots show the MSE-optimal grid sizes, the green lines represents $\hat{m}_{\mathsf{p}}$. The respective $RMSE$ is reported in brackets.

at a moderate level we only considered exchangeable copulas.[17]

[17]A copula is called exchangeable if $C(u_1, \ldots, u_d) = C(u_{\pi(1)}, \ldots u_{\pi(d)})$, for any permutation

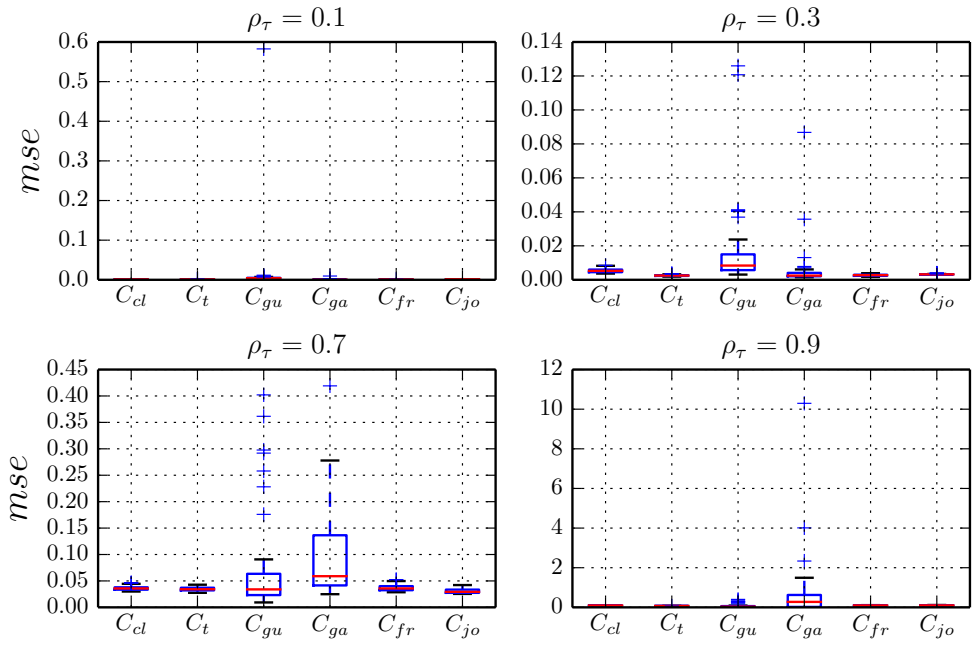

Figure 4.13.: Box plots of mean squared errors for selected levels of Kendall's tau in the 3-dimensional case.

4.3.6. Non-linear least squares estimate of $\hat{m}_{\mathsf{p}}$

Given the data generated by the monte carlo simulation we can take a slightly different approach and fit our model to the data with a least squares estimation procedure. We conduct this study to compare the parameterization we specified, namely choosing θ to be a rank correlation coefficient, with the parameter estimates $\hat{\theta}^{NLS}$ that are obtained from a least squares estimation. Furthermore, this will give us an indication of the goodness of fit of our model within a regression framework. We will not impose the restriction $\hat{\theta}^{NLS} \in [-1, 1]$ since we are explicitly interested in the natural parameterization of the model. We thus focus on the perspective of fitting a model to data that should capture the relationship between n and m_{opt}^{MSE} *without* a priori providing an interpretation of the regression coefficient θ^{NLS}. By comparing θ^{NLS} with the respective rank correlations we can analyze the optimality of our parameterization with respect to the situation that m_{opt} were an observable quantity. Since $\hat{m}_{\mathsf{p}}$ is non-linear in θ and an application of a logarithmic transformation would not succeed in linearizing our estimator with respect to this parameter,

$(\pi(1), \ldots, \pi(d))$ of (1, …, d). See e.g. McNeil et al., 2010.

we have to conduct a non-linear least squares estimation.[18] Thus we specify the following regression:

$$\mathbf{y} = |\boldsymbol{\theta}| e^{|\boldsymbol{\theta}|^{\frac{1}{\mathbf{n}}}} \mathbf{n}^{\frac{2}{d+4}} + \boldsymbol{\epsilon} \tag{4.39}$$

where

$$\mathbf{y} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{n} = \begin{pmatrix} n_1 \\ \vdots \\ n_n \end{pmatrix}, \boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_n \end{pmatrix} \text{ and } \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

.

Similar to linear regression, the least squares estimator of the unknown vector valued parameter $\boldsymbol{\theta} = (\theta_1, \ldots \theta_p)$ in the general model

$$y = f(\theta) + e,$$

where the errors e are assumed to be $i.i.d.$ with $\mu = 0$ and unknown variance σ^2, is the p by 1 vector $\hat{\theta}$ that minimizes

$$\| y - f(\theta) \|^2. \tag{4.40}$$

The estimate of the variance of the errors is given by

$$s^2 = \frac{\| y - f\left(\hat{\theta}\right) \|^2}{n - p}.$$

In non-linear least squares models the normal equations, that are obtained by taking the derivatives of (4.40) with respect to each θ_i and setting them to zero, are functions of the parameters, thus explicit solutions to this minimization problem cannot be obtained and one has to consider numerical methods. The results reported in the present analysis were obtained by applications of the Levenberg-Marquardt algorithm, a widely used algorithm in this field.

The penalized grid estimator $\hat{m}_{\mathsf{p}}$ we introduced performs promising in the dimensions we tested it. Apart from the advantages in precision it has compared to the least squares cross validation estimator m_{opt}^{lscv}, it proves to be significantly more robust and has clear advantages with respect to the computational cost of the calculation.

[18] The notation and general results with respect to non-linear regression in this section follow Gallant, 2009.

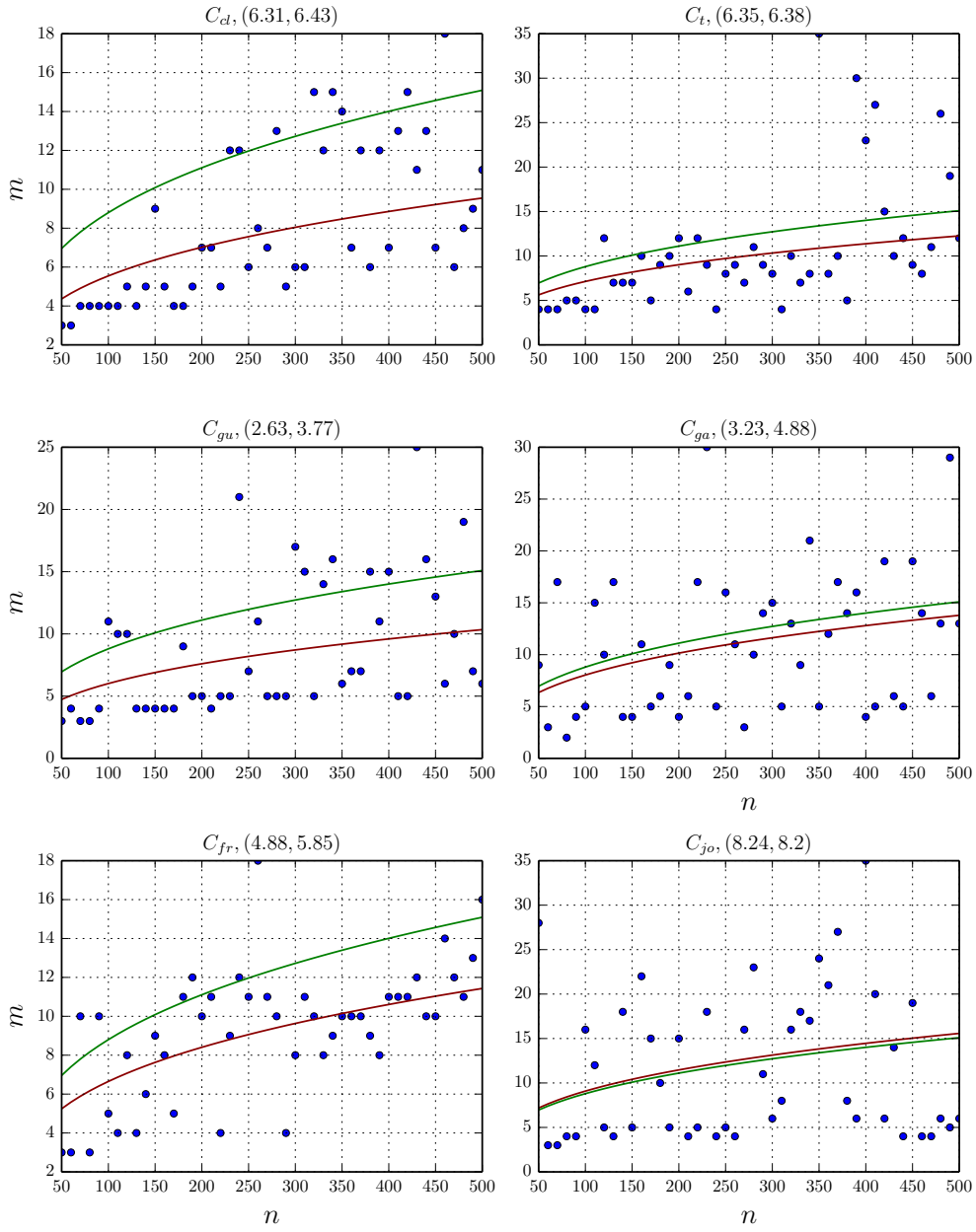

Figure 4.14.: Non-linear least squares estimates (red) of m_{opt} plotted against the MSE optimal grid sizes and $\hat{m}_{\mathsf{p}}$ (green) for different datasets. The $RMSE$ for m_{opt}^{NLS} and $\hat{m}_{\mathsf{p}}$ with respect to the data are reported in brackets as $(RMSE^{NLS}, RMSE^{p})$.

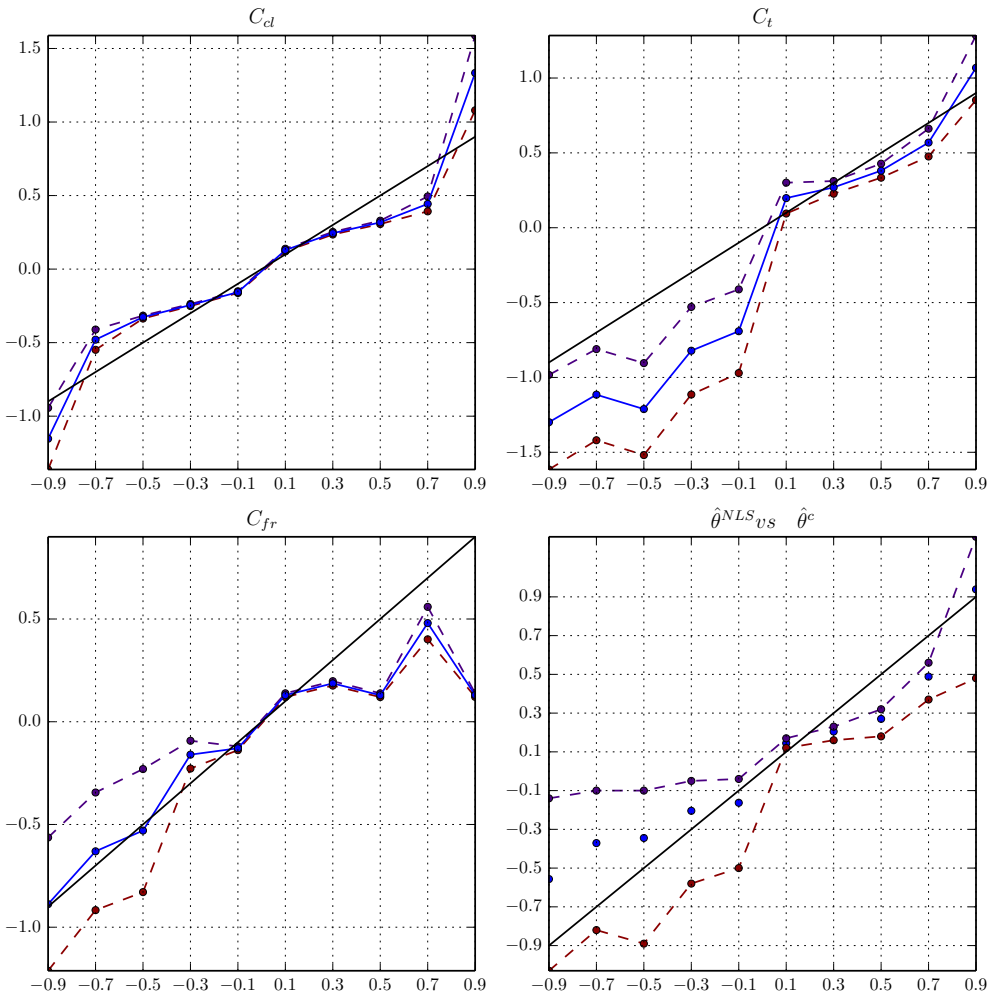

Figure 4.15.: Parameter estimates of NLS estimation with 2σ-intervals. The figure in the low right shows the mean across all samples with respect to the estimated parameters θ^{NLS}. Confidence intervals for the latter were calculated by bootstrapping[19] the parameter estimates for each copula at each value of Kendall's tau. The black line represents the parameters as specified by θ^p .

5. Locally Optimal Bernstein Copulas

In the previous chapter we introduced the estimator for an optimal grid size $\hat{m}_{\mathsf{p}}(\theta)$. This estimator leads to copula estimates that perform reasonably well in the center of the respective sample, but suffer from the well known drawback of non-parametric density estimators, that estimates in the tails of the respective distributions are poor for small to moderate sample sizes since data points in these regions are sparse. We can try to improve these estimates by utilizing dependence structure information that is local to the sample and estimate the density in the proximity of these points by Bernstein copula densities that are parameterized with an optimal grid size obtained by this additional information. Since we have given empirical evidence to the validity of the relationship between the MSE-optimal grid size and the strength of concordance, it is natural in our case to choose a suitable concordance measure ρ_ξ to represent this additional information. This leads to any given d-dimensional dataset $\mathbf{X_n^d} = (X_{1,i}, \dots X_{d,i}), i \in \{1, \dots n\}$, where n denotes the sample size, being representable by a family of locally optimal Bernstein copulas B given by:

$$B = B^d_{m(\theta)}, \forall\, \theta \in \Theta_{\mathbf{X}},$$

where the parameter-space $\Theta_{\mathbf{X}}$ is given by the set of locally realized concordance measures $\rho_{\xi,j}, j = 1, \dots, n$ at each point. But care has to be taken here, Boyer et al. (1997) have shown in the bivariate normal case that by calculating Pearson's correlation coefficient based on one variable being in some subset of the sample, one can arrive at wrong conclusions. The authors show that in this case for any subset $\mathcal{A}$ the correlation can be expressed as:

$$\hat{\rho}(\mathcal{A}) = \rho \left(\rho^2 + (1 - \rho^2) \frac{\mathrm{Var}(x)}{\mathrm{Var}(x) | x \in \mathcal{A}} \right)^{-\frac{1}{2}}$$

Note that in this case the conditional correlation is dependent on the ratio of the conditional variance of x to the overall variance in the sample. The larger the conditional variance is, the larger the conditional correlation will get *without* there being a change in the fundamental dependance relationship within the conditioning set.

Berentsen and Tjøstheim (2013) discuss various concepts of local dependence measures and point out that the conditional correlation coefficient, which can be

expressed as:

$$\hat{\rho}_c(\mathcal{A}) = \frac{\sum_{(X_{i,1},X_{i,2}\in\mathcal{A})} \left(X_{i,1} - \hat{\mu}_{X_{1,c}}\right)\left(X_{i,2} - \hat{\mu}_{X_{2,c}}\right)}{\sqrt{\sum_{(X_{i,1},X_{i,2}\in\mathcal{A})} \left(X_{i,1} - \hat{\mu}_{X_{1,c}}\right)^2}\sqrt{\sum_{(X_{i,1},X_{i,2}\in\mathcal{A})} \left(X_{i,2} - \hat{\mu}_{X_{2,c}}\right)^2}}$$

suffers from a drawback called the bias problem of conditional correlation which refers to conditional correlation failing to aggregate appropriately to the global Pearson's correlation coefficient which is incorrect for a pair of jointly Gaussian variables by definition. Within the research that analyzes financial contagion an expression for this bias has been developed relating the unconditional correlation to the correlation in times of crisis:[1][2]

$$\hat{\rho} = \frac{\rho_y}{\sqrt{1 + \left(\frac{\sigma^2_{y,1} - \sigma^2_{x,1}}{\sigma^2_{x,1}}\right)\left(1 - \rho_y^2\right)}},$$

where ρ_y denotes the conditional correlation in times of crisis and σ_x^2 denotes the conditional variance in regular periods. We observe the same deficiency as noted by Boyer et al. (1997), namely that the conditional correlation is prone to be biased due to changes in the variance of the conditioning set. Berentsen and Tjøstheim (2013) further point out, that using a linear dependence measure may not be appropriate in all cases and that conditional correlation is based on regions, which have to be specified in advance. They propose an alternative measure which they termed local Gaussian correlation and that is based on fitting Gaussian densities at each point of a bivariate sample by means of local likelihood estimation.[3] Since Pearson's correlation coefficient is sufficient for completely specifying the dependence structure of elliptical distributions, they thus completely describe the dependence structure in the neighborhood of the respective point under investigation.

Remembering the similarities between rank correlation and linear correlation, we thus have to be aware of potential inference problems that may occur due to the conditioning we will conduct in the following sections. We note that we are explicitly interested in the deviation of the local rank correlation coefficients from the unconditional rank correlation. The results obtained with respect to financial contagion analysis suggest that there is a relationship between the variances of the individual elements of the random vector and the pairwise rank correlations. The phenomenon of changes in the variances of a random variable is well studied in statistics and common methods of dealing with it include transforming the data or modeling the underlying functional relationship that drives these changes. We propose to utilize

[1] see e.g. Dungey* et al., 2005.
[2] see e.g. Forbes and Rigobon, 2002.
[3] see e.g. Loader, 1996.

the information whether Homoskedasticity is present or not to enhance our estimator via regionally measured rank correlations.
In contrast to the scope of the analyses conducted by the researchers, working in the field of financial contagion analysis, we are not trying to find evidence that the unconditional correlations for pairs of random variables have changed, we are trying to improve our estimation procedure with the usage of information that is inherently local. Under mild assumptions on the local smoothness of the underlying density we can justify this approach, which deviates from the holistic perspective of specifying the unconditional density. This method clearly has drawbacks since it relies on reducing the respective sample to smaller subsets, which offsets the idea of incorporating all information that is available for the analysis. The intuition, that lies behind this approach is that certain regions of a sample provide more information with regards to the local structure of a density than others. If one is interested in estimating the tail of a distribution where data is sparse, one should focus on specifying a full distribution in the neighborhood of the tail. The feasibility of this approach may be inferred from the fact, that every distribution can be represented as mixture of other distributions and analogous every copula can be represented as mixture of other copulas. An alternative path of local estimation that could be taken, which has the advantage of utilizing all the information that is provided by the dataset would be to apply a weighting function to the sample and estimate the Bernstein copula with respect to this weighted dataset. This approach clearly resembles kernel density estimation and the local likelihood procedure mentioned above and constitutes a hybrid method incorporating elements of these alternative non-parametric procedures. Analyzing the properties of this approach is subject to future research.

We now proceed to define the concept of weighted locally optimal Bernstein copulas, which will be our main model in this chapter.

Definition 1. *Let* $\mathbf{X_n^d} = (X_{1,i}, \dots X_{d,i}), i \in \{1, \dots n\}$ *be a d-dimensional sample, where n denotes the sample size. Let* $\Theta_{\mathbf{X}}$ *denote the set of locally realized concordance measures* $\rho^{\Omega}_{\xi,j}, j \in \{1, \dots n\}$ *with respect to the information set* Ω *in the sample and let* $\mathcal{B}_n = B^d_{m(\theta)}$ *be the family of locally optimal Bernstein copulas parameterized by* $m(\theta), \forall\, \theta \in \Theta_{\mathbf{X}}$.
Then

$$B^d_w = \sum_{i=1}^{n} \omega_i B^d_{m(\theta_i)}$$

is called a weighted locally optimal Bernstein copula, where $\omega = (\omega_1, \dots, \omega_n)$ *are weights satisfying* $\sum_1^n \omega_i = 1$.

Since every convex sum of copulas constitutes a copula[4], weighted locally optimal

[4]see Nelsen, 2007, chapter 3.2.4.

Bernstein copulas satisfy the necessary properties for them to be valid copulas. We will refer to the Bernstein copula resulting from the special case $\omega_i = 1, \omega_j = 0, \forall j \neq i$ as locally optimal Bernstein copula.

5.1. Specifying the information region

Within the models we develop in this chapter we are free to specify the regions we find relevant to our local information set Ω, i.e. the subsets of the sample $\mathbf{X}_n^d$ that we choose to use for the estimation of the local concordance measure $\rho_{\xi,i}, i \in \{1, \dots n\}$. It is theoretically possible to specify different information sets for every point in a given sample $\mathbf{X}$, if that is feasible for the analysis at hand. In the following we will discuss two specifications of Ω.

5.1.1. The spherical information set - $\Omega_\mathbf{s}$

Let $\boldsymbol{x}_i = (x_{i,1}, \dots, x_{i,d})$, $i \in \{1, \dots, n\}$ denote a point in d-dimensional euclidean space. The spherical information set around $\boldsymbol{x}_i$ is then constructed by evaluating Kendall's tau across all points of the d-ball with center $\boldsymbol{x}_i$ and radius r_i

$$\boldsymbol{\Omega}_\mathbf{s} := \{x \in \mathbb{R}^d : \|x - \boldsymbol{x}_i\|_2 \leq r_i\}, \tag{5.1}$$

where $\|\cdot\|_2$ denotes the euclidean norm. We allow r_i to vary across the whole sample for maximum flexibility. Depending on the specification of r_i local estimates for m_{opt} in the center of the sample may be close to the estimate m_{opt}^c gained on the whole dataset, whereas estimates taken in the other regions of the sample can be substantially different. Let $\mathbf{s}_n$ be the vector of sum-norms of the sample $\mathbf{X}$:

$$\boldsymbol{s}_n = \left(\| (x_1, \dots, x_d) \|_1^1, \dots, \| (x_1, \dots, x_d) \|_1^n\right)^\top.$$

Then

$$D_n = \max(\boldsymbol{s}_n) - \min(\boldsymbol{s}_n)$$

denotes the maximum distance between two points within the sample, i.e. in 2 dimensions that are farthest in the respective tails of the plane. We could then set $r = r_i = \frac{D_n}{2}, i = 1, \dots, n$ to construct spherical information regions, that cover at least 50% of the samples.

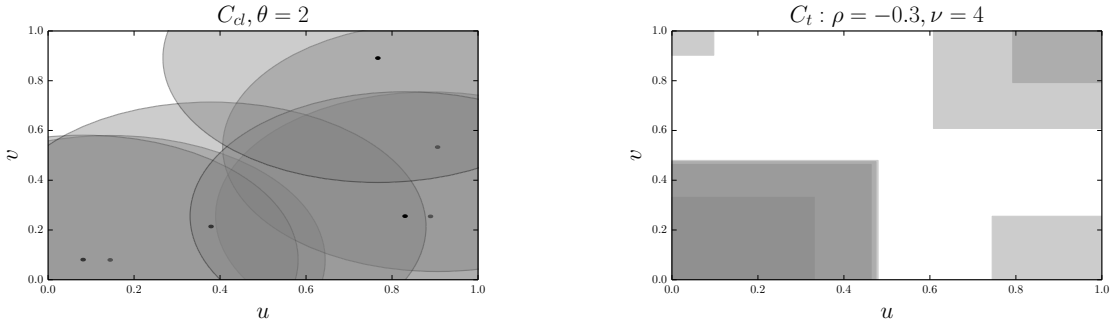

Figure 5.1.: Random samples drawn from a bivariate Clayton copula. The spherical information set is shown on the left, the quadrant-dependent information set on the right.

5.1.2. A quadrant-dependent information set - $\Omega_\mathbf{q}$

The second approach we consider is to split the sample into n-dimensional cubes. Let

$$Q = \{[0, 0.5]_1,]0.5, 1]_1\} \times \ldots, \times \{[0, 0.5]_d,]0.5, 1]_d\}$$

Then for every point $\boldsymbol{p}$ in $Q_j, j \in \{1, \ldots p\}$, where $p = 2^d$ denotes the number of quadrants, we can define a quadrant-dependent information set Ω_q by constructing a hypercube $\mathbf{C}_j(\boldsymbol{p})$ spanned at $\boldsymbol{p}$, with edge length $a = \|\boldsymbol{p}\|_\infty$ that lies within Q_j.

We can think of $\mathbf{C}_j(\boldsymbol{p})$ as a matrix, that is a scaled down version of Q_j, i.e. we arrive at $\mathbf{C}_j(\boldsymbol{p})$ by applying a suitable transformation $\mathbf{T}$ to the boundaries $\boldsymbol{b}_j$ of Q_j. The kth scaling factors $\boldsymbol{s}_j^k = (s_1, s_2), k = 1, \ldots, d$ of the hypercube are obtained by solving the set of equations

$$\boldsymbol{b}_j^k \mathbf{T} = \boldsymbol{a}, \quad k \in \{1, \ldots d\}.$$

where

$$\boldsymbol{b}_j^k \in \{(0, 0.5)_i^\top, (0.5, 1)_i^\top\},$$

$$\boldsymbol{a} = \begin{cases} (0, a)^\top, & \text{if } \boldsymbol{b}_j^k = (0, 0.5)^\top \\ (a, 1-a)^\top, & \text{if } \boldsymbol{b}_j^k = (0.5, 1)^\top \end{cases},$$

and

$$\mathbf{T} = \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix}.$$

The kth row-vector of $\mathbf{C}_j(\boldsymbol{p})$ is then obtained by:

$$\boldsymbol{c}_j^k(\boldsymbol{p}) = \boldsymbol{s}_j^k \circ \boldsymbol{b}_j^k = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \circ \begin{pmatrix} b_{1,j}^k \\ b_{2,j}^k \end{pmatrix},$$

where $\circ$ denotes the Hadamard product. We can then construct Ω_q with respect to to the point $\boldsymbol{p}$ by considering all points within the boundaries given by the Cartesian product of the rows of $\mathbf{C}_j(\boldsymbol{p})$.

The construction of Ω_q is driven by the aim to provide a mechanism that automatically adapts to the tail-behavior of a dataset, and to incorporate this information appropriately to the estimation of $\hat{m}_{\mathsf{p}}$. Whereas Ω_r can be regarded as utilizing information that is local but still symmetric at each point and should thus differ less from $\hat{m}_{\mathsf{p}}$.

Figure 5.2 shows the regions of varying $\hat{m}_{\mathsf{p}}(\theta)$. We can observe, that the quadrant dependent information set leads to higher values of the estimator in the respective tails, than the grid sizes based on the spherical information set.

5.2. Time Varying Bernstein Copulas

The LBC framework is easily extended to timeseries applications and time-dependent dependence structures, by specifying the information set appropriately:

$$B^d_{m(\theta_t)} = B^d_{m(\theta_t)} | B^d_{m(\theta_{t-1})} \dots B^d_{m(\theta_T)}, \forall\, \theta \in \Theta, t \in \{1 \dots T\}, \tag{5.2}$$

where the index t denotes a point in time. This specification has clear parallels to conditional copula models as introduced by Patton (2006) with the difference that we are adjusting the bandwidth that is needed to estimate the underlying copula conditioned on historical values of θ.

Before we present an application of the time varying LBC, it is important to take a clear perspective what type of dependency one wishes to address with the respective copula model. Consider a d-dimensional timeseries $\mathbf{X} = (X_1 \dots X_d)_t$. We can analyze this series with respect to its cross-sectional dependence structure, i.e. the dependency between the d variables at time t, or we can analyze each of the d series with regards to their inter-temporal dependence structure, i.e. in the bivariate case the dependence structure between $X_{i,t}$ and $X_{i,t-j}$, $j \in \{1, \dots, t-1\}$.[5]. Studies of the former are conducted by analyzing the dependence structure of the residuals of the d timeseries, where one is either interested in the unconditional dependence structure, i.e. the dependence structure C of the $i.i.d$ sample of random vectors

[5] see e.g. Patton, 2009, for a review of the literature on the various aspects of time varying copula analysis with regards to financial time series.

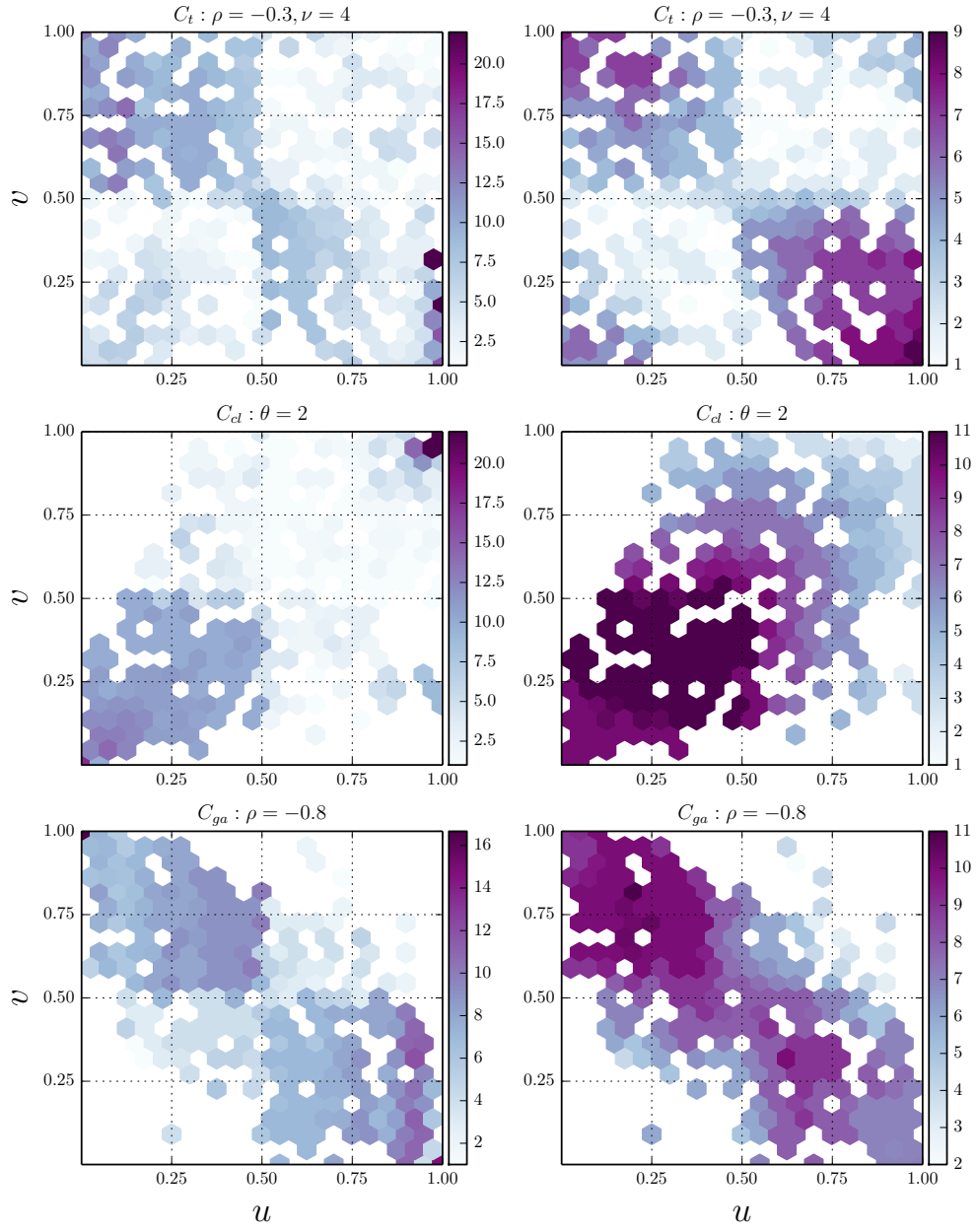

Figure 5.2.: Scatter Plots of regions of local optimal grid sizes for different copulas in the bivariate case. Grid sizes generated by Ω_q are shown on the left, grid sizes based on Ω_s set are shown on the (right). The color-spectrum indicates the values of $\hat{m}_{\mathsf{p}}(\theta)$ in the respective regions.

$\boldsymbol{\varepsilon}\,(\epsilon_1, \ldots, \epsilon_d)_t$ or one focuses on the conditional dependence structure $C^t_{\nu,\rho}|\mathcal{F}_{t-j}$ of the random vectors $\boldsymbol{\varepsilon}|\mathcal{F}_{t-j} = (\varepsilon_1|\mathcal{F}_{t-j}, \ldots \varepsilon_d|\mathcal{F}_{t-j})_t$.

When parameterizing the Bernstein copula according to Equation (5.2) it is beneficial to remember that by adjusting $\hat{m}_{\mathsf{p}}$ according to different information sets one poses no restriction on the estimand, i.e. we are not necessarily estimating a possibly different copula with a different parameter. We are indeed inferring the structure of the copula solely from the data and are adjusting the grid size to the optimality criterion specified in the previous chapter. If there are structural breaks or other changes in the dependence structure of the underlying model the LBC will automatically adjust accordingly *if* the change can be captured by the concordance measure or the Bernstein copula. The possibility of specifying different information sets allows the LBC to adjust even more flexibly to asymmetries in the data than the regular Bernstein copula. Jumps in the concordance measure for a given sample size can then be interpreted as additional information that lets the data being more or less reliable, therefore leading to increased or decreased smoothing.

5.2.1. Application: Forecasting financial data

The applications of copula-models in finance are numerous. They are used in quantitative risk management e.g. to model the aggregate dependence structures of financial conglomerates by coupling the risks that are represented by the conglomerates subsidiaries. On this aggregate basis one can then evaluate various downside risk measures such as the Value at Risk or the Expected Shortfall of the conglomerate. Calculations of the latter form are used for stress testing purposes and are proprietary for financial institutions in Europe due to the regulatory frameworks of Basel 2 and Solvency 2, if they exceed certain criteria related to their size. A conglomerate that decides to use a so called internal model can choose to use copula models to determine their overall risk exposure. Companies will do so if they recognize that the alternative, that is represented by having to apply the so called standard formula, places inappropriate capital requirements on them. Another field where copulas are widely used is option pricing and the pricing of credit derivatives, such as Collateralized Debt Obligations. To derive fair values in the latter case one often has to specify the dependency structure of multiple underlyings/obligors since the respective products constitute cash flows that are directly related to the probabilities of certain events happening.[6] In the following we will apply the LBC framework to analyze the time varying dependence structure of the S&P 500, Nikkei and Dax and use the resulting model to forecast their future dependency structure. This analysis may be interesting in the present times as questions regarding the stability of international financial markets with respect to their interdependencies have become prominent after the financial crisis in 2007. Quantifying the dependance structure between the leading stock market indices of three major economies may

[6] see e.g. Cherubini et al., 2004.

give a first insight into the dynamics and the structure of intercontinental financial relationships.

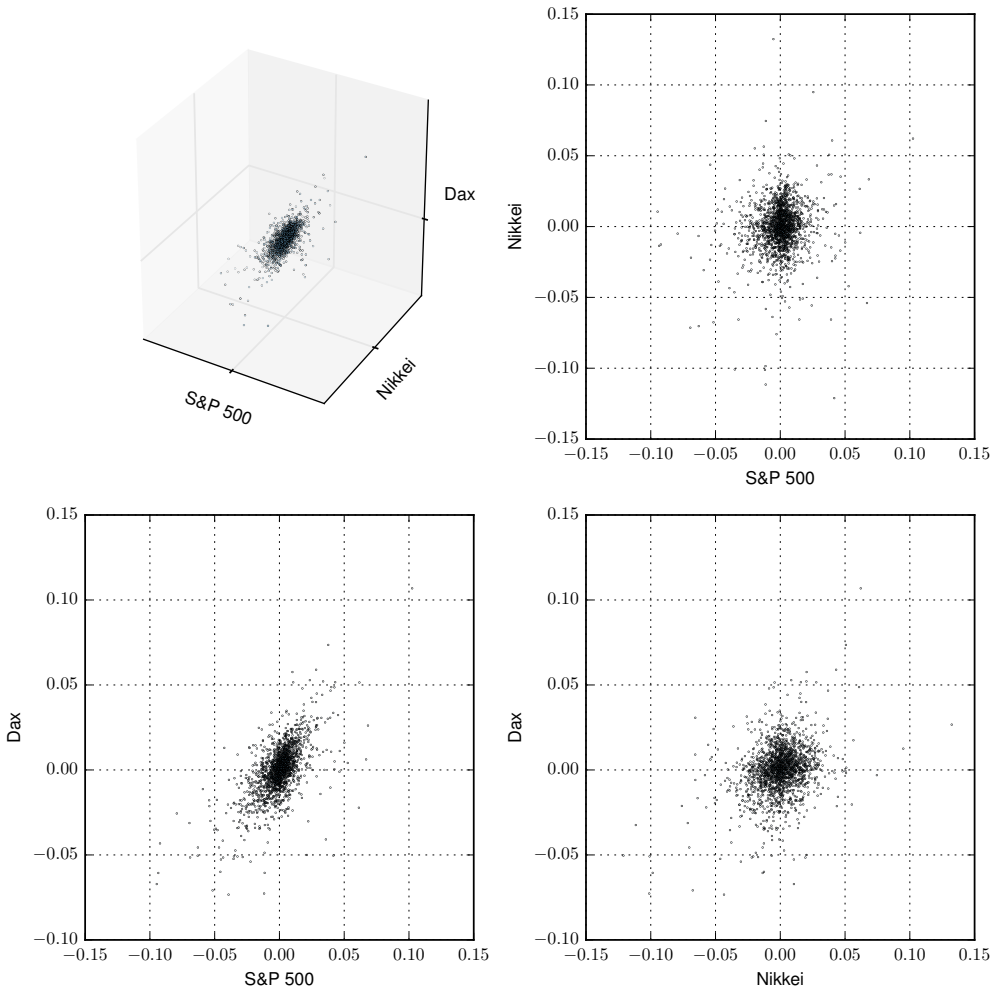

Figure 5.3.: Scatter Plots of stock market indices.

The time-varying LBC-model is operationalized by specifying a full model with respect to the dynamical properties of the marginal distributions and the concordance measure that parameterizes the LBC. Estimating this model is therefore performed in a two-step procedure. The first step consists of estimating the marginal processes and the process used to quantify the dynamics of the parameter, whereas the second step couples the margins with the Bernstein copula, parameterized by $\hat{m}^t_{\mathsf{p}}$. A forecast within this model then consists of forecasting the future value of the parameter and then calculating the Bernstein copula density associated with this parameterization. Models like these have been applied in parametric settings by e.g. Dias

and Embrechts (2004), Jondeau and Rockinger (2006) and Patton (2006). To our knowledge their hasn't been an application within a non-parametric framework yet.

Table (5.1) summarizes descriptive statistics for the return series of the three indices. We choose to consider 1832 joint observations within the period from January 1, 2007 up to October 3, 2010. The adjusted closing prices of the indices were transformed to logarithmic returns, as is standard in quantitative finance due to the various mathematical benefits with respect to interpretation, normalization, stabilization and mathematical ease of the transformed series. The mean, median and the quantiles of each series indicate that the respective distributions can be considered symmetrical. Since we are considering stock market indices we naturally have to consider GARCH-models for modeling the return data since these timeseries usually show clear signs of time-varying volatility clustering.

Table 5.1.: Descriptive statistics for the returns of the three stock market indices

	Dax	Nikkei	S&P 500
count	1832.00	1832.00	1832.00
$\hat{\mu}$	0.00	-0.00	-0.00
$\hat{\sigma}$	0.01	0.02	0.01
min	-0.07	-0.12	-0.09
q_{25}	-0.01	-0.01	-0.00
$q_{0,5}$	0.00	0.00	0.00
$q_{0.75}$	0.01	0.01	0.01
max	0.11	0.13	0.10

Modelling the margins

This assumption is confirmed graphically by inspecting Figure 5.4 which depicts various plots for the respective indices. We can observe the typical characteristics of financial timeseries, i.e. volatility clustering in the return series and persistent patterns of autocorrelation in the squared return series.

The Ljung-Box Test for autocorrelation in the squared return series depicted in Table 5.2 confirms the findings that we graphically observed in the ACF and PAC-plots, as we can reject the null hypotheses of no autocorrelation for every series and every lag.

According to these observations we consequently choose to specify GARCH(p, q) models for all indices, where we set $p = q = 1$ since these perform traditionally well

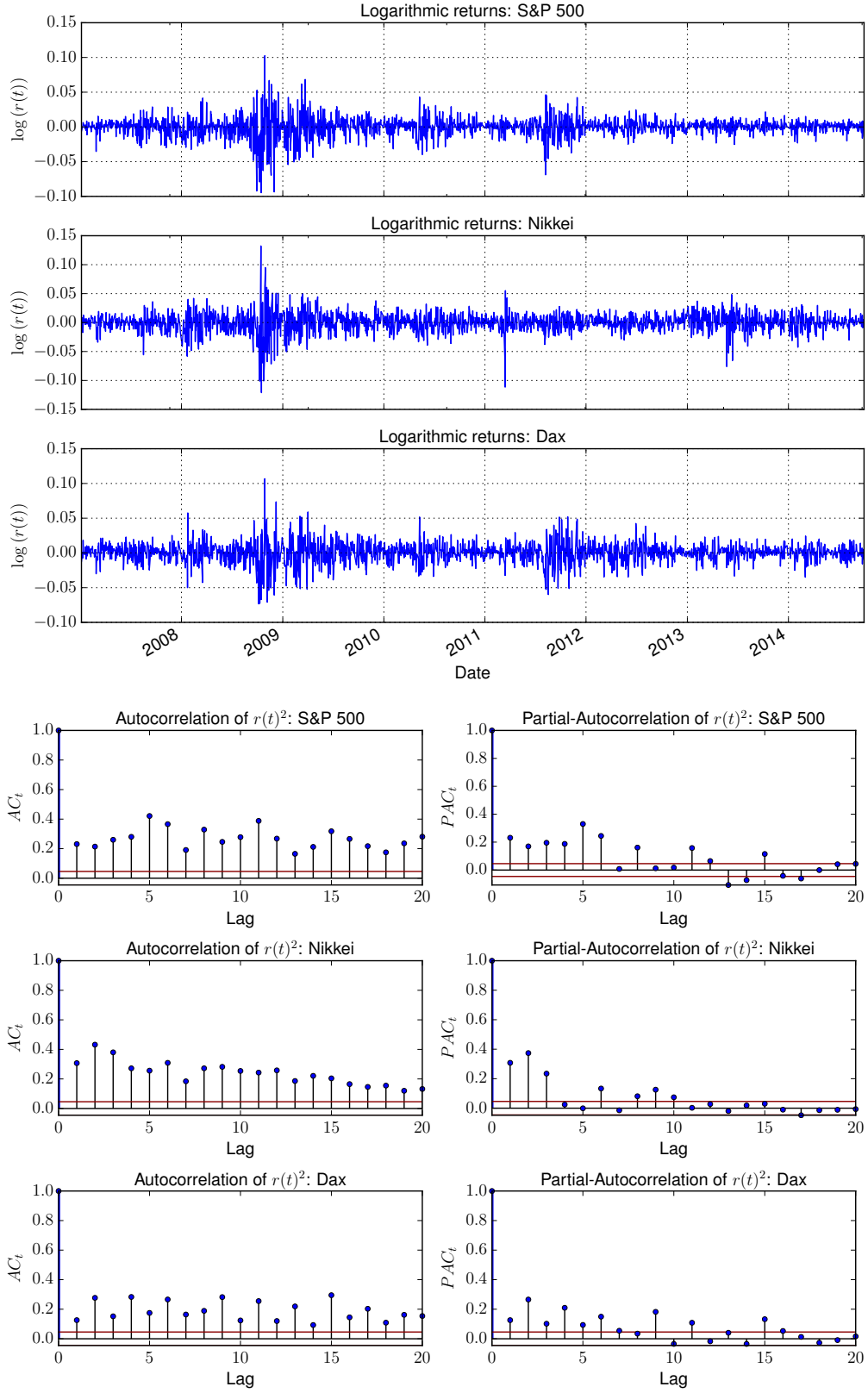

Figure 5.4.: Volatility clustering, autocorrelation and partial-autocorrelation of squared returns for the S&P 500, Nikkei and Dax respectively.

Table 5.2.: Ljung-Box Test for autocorrelation in the squared returns

		Dax	Nikkei	S&P 500
Lag 10	χ^2	835.474	1676.188	1543.504
	p-value	0	0	0
Lag 15	χ^2	1247.375	2138.865	2274.022
	p-value	0	0	0
Lag 20	χ^2	1477.619	2332.184	2798.273
	p-value	0	0	0

on financial timeseries, that show the characteristics mentioned above:

$$r_t = \mu + \epsilon_t \,, \tag{5.3}$$
$$\epsilon_t = \sigma_t z_t \,, \tag{5.4}$$
$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \tag{5.5}$$

In addition to this GARCH(1,1) model we estimated two ARMA-GARCH models with GARCH-in-mean (GARCH-M) specifications which allow for various dynamics with respect to to the mean-model depicted in Equation (5.5), as well as providing the ability to incorporate lagged variables of both returns and innovations:

$$\Phi(L)(r_t) = \mu_t + \Theta(L)\epsilon_t \,, \tag{5.6}$$

where L denotes the *lag operator*. The dynamics for μ_t are given by

$$\mu_t = \mu + \theta f(\sigma). \tag{5.7}$$

We set $\mu \in \{0, \mu\}$ and $f(\sigma) = \sigma$ in Equation (5.7), i.e. we specify the following two GARCH-M(1,1) models :

$$r_t = \mu_t + \epsilon_t \,, \tag{5.8}$$
$$\mu_t = \theta\sigma \,, \tag{5.9}$$
$$\epsilon_t = \sigma_t z_t \,, \tag{5.10}$$
$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \,, \tag{5.11}$$

which we will denote GARCH-M(1,1)$_1$ and

$$r_t = \mu_t + \phi r_{t-1} + \epsilon_t + \psi\epsilon_{t-1}, \tag{5.12}$$
$$\mu_t = \mu + \theta\sigma, \tag{5.13}$$
$$\epsilon_t = \sigma_t z_t, \tag{5.14}$$
$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{5.15}$$

which we will denote GARCH-M(1,1)$_2$.
We can confirm that these model specifications are appropriate by analyzing the autocorrelation functions of the standardized squared residuals, which should show no signs of autocorrelation, if the model that generated the residuals is able to replicate the characteristics of the underlying time series. For completeness we thus report Ljung-Box Tests for autocorrelation in the squared residuals of the estimated GARCH-M(1,1)$_2$ series in Table 5.3. The results for the other specified processes are similar and given in Appendix A.3. Let Ω_t denote $\{r_t, r_{t-1}, ...\}$. To model the

Table 5.3.: GARCH-M(1,1): Ljung-Box Test for autocorrelation in the standardized squared residuals

		Dax	Nikkei	S&P 500
Lag 10	χ^2	15.15	10.09	3.95
	p-value	0 .127	0.432	0.950
Lag 15	χ^2	16.373	13.922	6.145
	p-value	0.358	0.531	0.977
Lag 20	χ^2	20.656	16.777	9.62
	p-value	0.418	0.667	0.975

conditional distribution of the innovations $\epsilon|\Omega_{t-1}$ we choose to specify t-distributions to account for the leptokurtic properties of the empirical densities as depicted in the histograms in Figure 5.5:

$$\epsilon|\Omega_{t-1} \sim t_\nu(0, 1), \tag{5.16}$$

where ν denotes the degrees of freedom of the respective t-distribution. Specifying the appropriate distribution for the innovations is important due to various reasons, as GARCH-models of financial timeseries may be used for pricing purposes or to measure the downside risk of the respective series. McNeil and Frey, 2000 use extreme value theory to estimate the tail of the innovation distribution and show that utilizing heavy-tailed distributions provide one-day estimates that are superior to the estimates that result from using distributions that neglect this property.

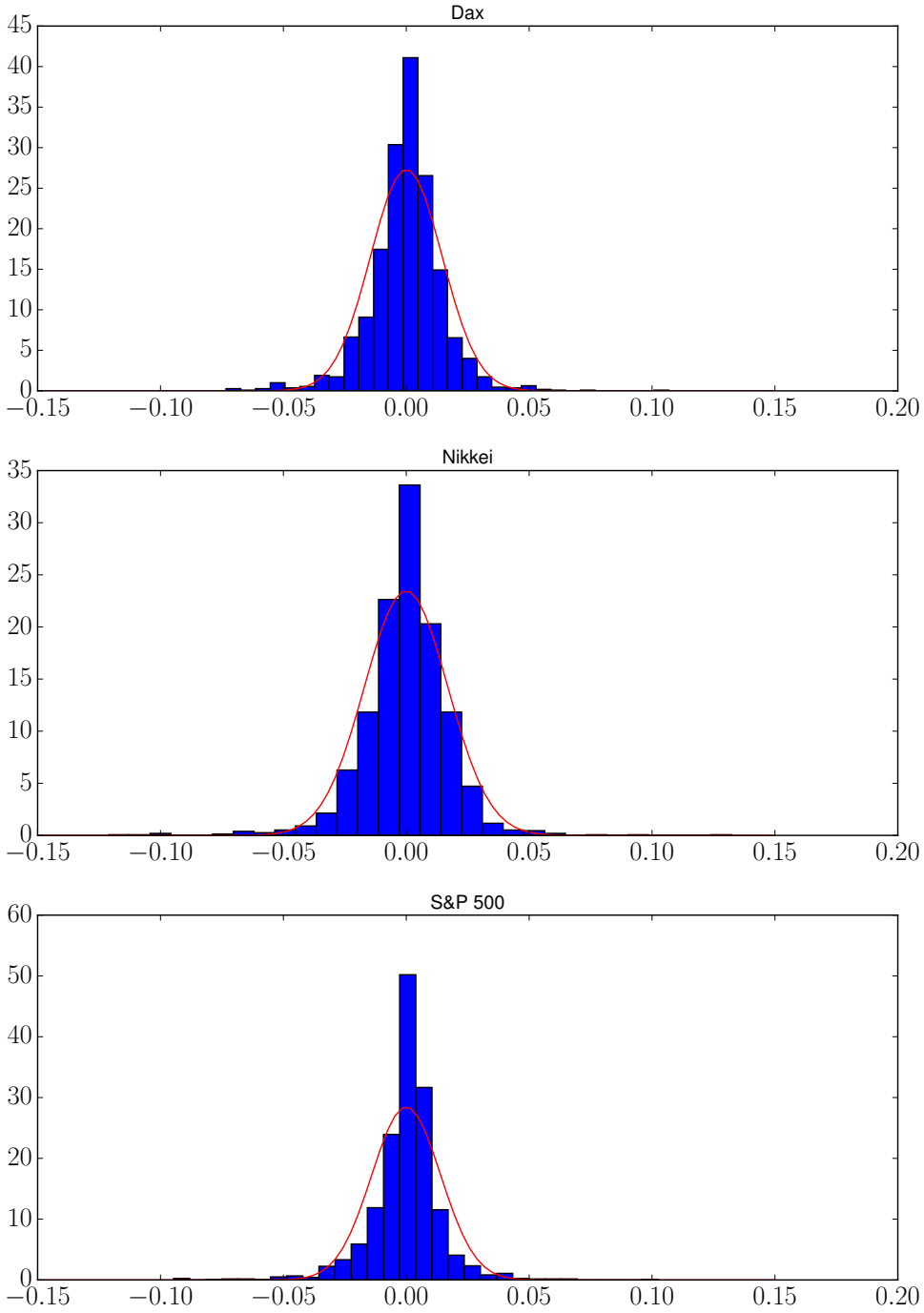

Figure 5.5.: Leptokurtosis in index data

Modelling the ρ_τ-Series

When modeling the ρ_τ-Series it is a priori not clear which time-series model might be appropriate, in the sense that there is no clear intuition as tho which model class might be appropriate. The dynamic behavior of rank correlation matrices might be infered from the dynamic behavior of correlation matrices that are constructed by dynamic conditional correlation models as developed by Engle, 2002 . This holds true with respect to Kendall's Tau for multivariate models that follow an elliptical distribution since we know the functional relationship between Pearsons correlation coefficient and Kendall's Tau. Nevertheless, in general such inference is not feasible, thus we have to consider a larger set of timeseries models in the specification step. In a first step, the estimation of an appropriate model requires us to select an appropriate process from the rank-correlation matrix $\rho_\xi(r)$. Following the arguments provided in Chapter 4.3.5, we choose to to identify the relevant timeseries according to the max criterion and for the reminder of this thesis we will denote $\rho_\tau^{\max}= \max|\rho_\tau|$. We generated the series by applying a moving window to the margins, where the windowsize represents the sample size that is used to calculate Kendall's Tau. The window w is moved according to a stepsize s that has to be specified carefully, as choosing a step size that is too small can lead to artificially highly correlated data.[7]

Dias and Embrechts, 2004 consider a GARCH-type specification to capture the dynamics of the underlying e-dimensional dependency parameter vector θ_t:

$$\theta_{m,t} = r_0 + \sum_{i=1}^{r} r_i \prod_{j=1}^{d} Z_{j,t-i} + \sum_{k=1}^{s} s_k \theta_{m,t-k}$$

where $m = 1, 2 \ldots, e$ and $r_i, i = 0, \ldots, r$ and $s_j = 1, 2, \ldots, s$ are scalar model parameters. Interestingly we found this GARCH-type specification to be inappropriate for our purposes, as modeling the $\rho_\tau^{\max}$-series with a ARMA-process provided the clearly superior results. One might provide a reason for this observation by inspecting the graphics depicted in Figure (5.6). The two graphics at the top show the $\rho_\tau^{\max}$-Series and the Fisher-transformation of the same series. The Fisher-transformation for ρ is known to induce normality in distribution and to be variance stabilizing, both under certain conditions, i.e. when the random variables under consideration follow a bivariate normal distribution and the sample size is large. If non-normality is present, then this transformation provides poor results with respect to these properties.[8] $\rho_\tau^{\max}$is bounded by 0 and 1, so we observe no realizations beyond these bounds,

[7] see Mittnik, 2011, The author reveals flaws within the calibration of correlation matrices that are provided to insurance companies that use the so called standard formula to calculate their regulatory capital requirements within the Solvency 2 regulatory framework.

[8] see eg. Berry and Mielke Jr, 2000.

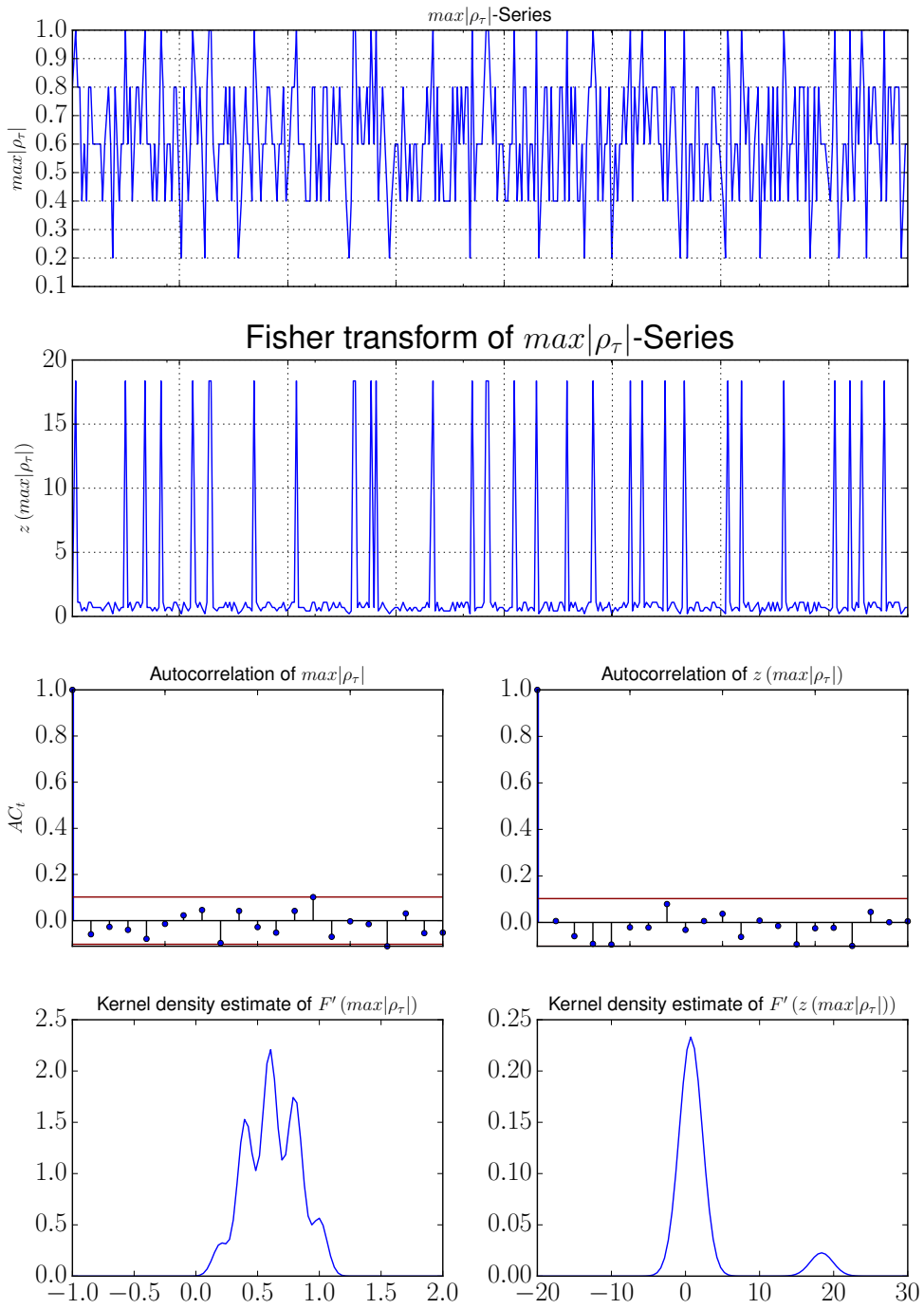

Figure 5.6.: Characteristics of $|\rho_\tau|$-Series

which leads to further questions regarding the distributional properties of this series. To assess the latter we estimate the densities of both series, the original and the transformed series, by the kernel density estimator, the graphics are depicted in the bottom row of Figure (5.6). We see that the Fisher-transformation still *stabilizes* the distribution, transforming the multi-modal distribution of the original series to a smooth bi-modal distribution. For inference purposes it is still advised to utilize bootstrapping to establish appropriate confidence intervals. From the autocorrelation functions depicted in the third row of (5.6) we can see that the series can be regarded as being i.i.d. which indicates that our choices for the step and window size, $(s = 5, w = 5)$ of the moving window were appropriate. Within the LBC framework the region we specified is identified by all data points that constitute a period of 5 consecutive days.

Table 5.4.: Descriptive statistics for the $\rho_\tau^{\max}$-series

	$\rho_\tau^{\max}$
count	366
$\hat{\mu}$	0.62
$\hat{\sigma}$	0.21
min	0.2
$q_{0.25}$	0.4
$q_{0.5}$	0.6
$q_{0.75}$	0.8
max	1.0

Table 5.4 summarizes descriptive statistics for the $\rho_\tau^{\max}$-series. The location of the mean, 25% and 75%-quantile again suggest a rather symmetrical distribution, which we can confirm graphically by the aforementioned kernel-density estimate depicted in the last row of Figure 5.6. The multiple peaks of the density function are only recognized visually, whereas the symmetrical nature of the distribution is revealed when focusing on the descriptive statistics. The probability mass that is appointed to values that a larger than the upper boundary of the theoretical support of Kendall's Tau are a result of the commonly known boundary bias problem of kernel density estimation. It is important to remember that these statistics are attributes of a very specific process, namely that of the maximum of the rank correlation matrices that are calculated on 5-day datasets, which are obtained from our index-series. The multiple occurrence of the upper boundary, depicted in Figure 5.7, is a result of these calculations on small datasets. We can indeed observe almost linear relationships between the respective indices, Dax and S&P 500, during these periods. As a result of our observations with regard to the autocorrelation functions we choose to

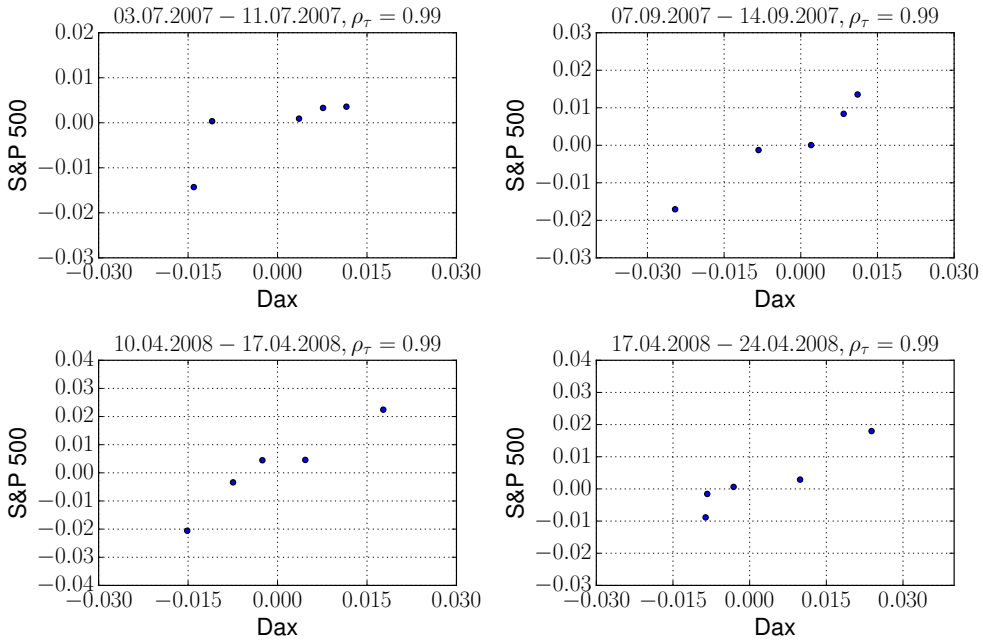

Figure 5.7.: Scatter plots of weekly subsamples with ρ_τ near 1.

estimate following ARMA(1,1)-model:

$$\rho_{\tau,t} = \mu + \phi\rho_{\tau,t-1} + \epsilon_t + \theta\epsilon_{t-1} \tag{5.17}$$

Parameter estimation

The results of the estimations of the GARCH model parameters are depicted in Tables 5.5, 5.6 and 5.7. For the GARCH(1,1) and GARCH-M$(1,1)_1$ models all estimated parameters are statistically significant for all three indices. The intercept for the σ_t^2-process is estimated to be 0 for both models and the α and β parameters are of similar magnitude for all indices.

The estimated shape parameter ν of the t-distribution of the innovations varies across all indices, with the highest value being estimated for the Nikkei-series, followed by Dax and S&P 500. This pattern is shared across the GARCH(1,1) and GARCH-M$(1,1)_1$-models, with the difference being that the estimated $\nu's$ are higher for the GARCH-M$(1,1)_1$-model. The interpretation of the varying shape parameters is straightforward, since as $\nu \downarrow 1$ the tails of the t-distribution become increasingly heavy and the marginals become less peaked about 0.[9] The results for the GARCH-M$(1,1)_2$-model are similar to the results of the other two models with re-

[9] see e.g. Kotz and Nadarajah, 2004.

Table 5.5.: GARCH(1,1): Parameter estimates for Dax, Nikkei and S&P 500

Dax				
	Estimate	Std. Error	t-value	p-value
μ	0.0007	0.0002	3.0366	0.0024
ω	0.0000	0.0000	2.6514	0.0080
α	0.1050	0.0170	6.1647	0.0000
β	0.8883	0.0164	54.0794	0.0000
ν	6.8825	1.1635	5.9155	0.0000
Nikkei				
	Estimate	Std. Error	t-value	p-value
μ	0.0005	0.0003	1.8856	0.0593
ω	0.0000	0.0000	2.9615	0.0031
α	0.1032	0.0170	6.0686	0.0000
β	0.8785	0.0192	45.8699	0.0000
ν	10.0420	2.1845	4.5970	0.0000
S&P 500				
	Estimate	Std. Error	t-value	p-value
μ	0.0009	0.0002	4.7287	0.0000
ω	0.0000	0.0000	2.9118	0.0036
α	0.1169	0.0187	6.2542	0.0000
β	0.8821	0.0164	53.6610	0.0000
ν	4.8024	0.6265	7.6655	0.0000

spect to the parameters of the σ_t^2-process, as all of these parameters share a similar magnitude as well as all being statistically significant. In addition to that the parameters of the mean-model are all significant for the S&P 500-series, whereas ϕ and ψ are insignificant for both, the Nikkei and the Dax-series. We computed the AIC and BIC information criterion to compare the specified models with each other and identify the most suitable model for each index-series. Table 5.8 summarizes the results for both criteria. Both criteria jointly favor the GARCH(1,1)-model for the Dax-series and the GARCH-M$(1,1)_2$-model for the S&P 500 series. The results for the Nikkei-series are less clear since AIC favors the GARCH-M$(1,1)_2$-model whereas BIC favours the GARCH-M$(1,1)_1$-model. We choose to give AIC precedence over BIC for the Nikkei-series and consequently choose the GARCH(1,1)-model for the Dax-series and the GARCH-M$(1,1)_2$ model for the Nikkei and S&P 500. The param-

Table 5.6.: GARCH-M(1,1)$_1$: Parameter estimates for Dax, Nikkei and S&P 500

Dax				
	Estimate	Std. Error	t-value	p-value
ω	0.0000	0.0000	2.6434	0.0082
α	0.0989	0.0160	6.1833	0.0000
β	0.8936	0.0157	57.0337	0.0000
ν	7.1994	1.2720	5.6597	0.0000
Nikkei				
	Estimate	Std. Error	t-value	p-value
ω	0.0000	0.0000	2.9548	0.0031
α	0.1016	0.0168	6.0526	0.0000
β	0.8798	0.0191	46.1400	0.0000
ν	10.3108	2.2949	4.4929	0.0000
S&P 500				
	Estimate	Std. Error	t-value	p-value
ω	0.0000	0.0000	2.9179	0.0035
α	0.1097	0.0174	6.2926	0.0000
β	0.8866	0.0159	55.7692	0.0000
ν	5.3307	0.7582	7.0311	0.0000

eters for the $max|\rho_\tau|$-process, which is based on the GARCH-processes we specified are reported in Table 5.9.

Forecasting

We now turn to forecasting the future value of $\rho_\tau^{\max}$, which is given by the h-step ahead forecast of the process in Equation 5.17, parameterized with the parameter values we have just estimated:

$$\hat{\rho}_{\tau,t+h} = \phi\hat{\rho}_{\tau,t+h-1} + \psi\hat{\epsilon}_{t+h-1}\,, h = 1, 2\ldots. \tag{5.18}$$

Table 5.7.: GARCH-M$(1,1)_2$: Parameter estimates for Dax, Nikkei and S&P 500

Dax				
	Estimate	Std. Error	t-value	p-value
μ	0.0007	0.0002	3.0386	0.0024
ϕ	-0.2723	0.5832	-0.4670	0.6405
ψ	0.2901	0.5799	0.5003	0.6168
ω	0.0000	0.0000	2.6536	0.0080
α	0.1049	0.0171	6.1372	0.0000
β	0.8884	0.0165	53.9324	0.0000
ν	6.8584	1.1585	5.9201	0.0000
Nikkei				
	Estimate	Std. Error	t-value	p-value
μ	0.0006	0.0003	1.9637	0.0496
ϕ	-0.3701	0.3708	-0.9983	0.3181
ψ	0.3237	0.3771	0.8584	0.3907
ω	0.0000	0.0000	2.9554	0.0031
α	0.1031	0.0170	6.0491	0.0000
β	0.8786	0.0192	45.7594	0.0000
ν	9.9020	2.1446	4.6172	0.0000
S&P 500				
	Estimate	Std. Error	t-value	p-value
μ	0.0009	0.0001	6.5568	0.0000
ϕ	0.7974	0.0878	9.0800	0.0000
ψ	-0.8551	0.0756	-11.3097	0.0000
ω	0.0000	0.0000	2.9266	0.0034
α	0.1179	0.0187	6.2978	0.0000
β	0.8811	0.0165	53.4349	0.0000
ν	4.7063	0.5968	7.8864	0.0000

Given the prediction for $\rho_\tau^{\max}$we define the Bernstein copula density estimator for period h as

$$c^d_{m\left(\hat{\rho}_{\tau,t+h}\right)} = \prod_{i=1}^{d} m_i \sum_{v_1} \dots \sum_{v_d} \alpha\left(\frac{v_1}{m_1} \dots, \frac{v_d}{m_d}\right) \prod_{i=1}^{d} P_{m_i,v_i}(u_i), \tag{5.19}$$

Table 5.8.: *AIC* and *BIC* for GARCH(1,1), GARCH-M$(1,1)_1$, and GARCH-M$(1,1)_2$

	GARCH(1,1)	GARCH-M$(1,1)_1$	GARCH-M$(1,1)_2$
		Dax	
AIC	-5.9525	-5. 9485	-5.9507
BIC	-5.9375	-5.9364	-5.9296
		Nikkei	
AIC	-5.6750	-5.6742	-5.6752
BIC	-5.66	-5.6621	-5.6541
		S&P 500	
AIC	-6.2895	-6.2788	-6.2967
BIC	-6.2744	-6.2667	-6.2756

Table 5.9.: Parameter estimates for $\rho_\tau^{\max}$-process

	Estimate	Standard Error	p-value
μ	0.8750	0.0872	0
ϕ	-0.9418	0.0635	0
θ	0.6180	0.0051	0

where $u_i = F_i(\epsilon_i)$ denote the cross sectional residuals, i.e. we parameterize the future Bernstein copula density by the residuals of the total sample, which are transformed to the unit interval by their estimated marginal distributions. We use this Bernstein copula to draw samples from the estimated future distribution of the stock market indices. These samples can be used to query certain aspects of interest with regards to the joint occurrence of these random variables.

Calculating Value-at-Risk and expected shortfall

To provide an application that operationalizes the time varying LBC we give an example from financial risk management and calculate Value-at-Risk (VaR) and expected shortfall (ES) estimates for a portfolio, which is constructed from future contracts that are based on the S&P 500, Nikkei and the Dax. We cannot invest directly into these indices as they represent a weighted average aimed at measuring the performance of key companies in the respected countries. To invest into one of these indices one has to apply either derivative strategies or invest in so called Exchange

Traded Funds (ETF's). Examples of the former are investments in certificates or futures, where the price of both of these instruments rely on the development of the underlying index. An Index ETF tries to replicate the performance of a specific index by either reconstructing all the components of the respective index or by investing in a significant sub-sample thereof. Certificates are structured products which represent payments that are contingent on the underlying and therefore share a lot of the characteristics of the wider known options. The owner pays a principal which can be augmented or reduced depending on the development of the underlying. Future-contracts are derivative products that specify a future exchange of the underlying at a fixed price, which is settled in the time of agreement.
Figure 5.8 depicts the index series plotted against their correspondent future series. The future series are scaled by 0.8 to emphasize the nearly identical paths of these derivatives. Except for the pair for the S&P 500 the paths of the respective series are nearly indistinguishable. The reason for this can be derived from an arbitrage argument, that is used to establish the fair price of the future. Since the future contract at time t, $F(t,T)$, fixes the price for an exchange of the commodity in the future $t = T$ it has to be priced at the spot price of the underlying compounded at the risk free rate r, i.e.

$$F(t,T) = S_t(1+r)^T, \tag{5.20}$$

as a different price would lead to arbitrage opportunities. This is easily seen when comparing a direct investment in the underlying with a long position in the future contract. Both strategies lead to the possession of the underlying in $t = T$ without taking any risk therefore they are perfectly equivalent and theoretically must have the same value at time t. Arbitrage arguments are in essence equilibrium arguments, i.e. they don't imply, that there are no opportunities for arbitrage, they imply that in an equilibrium in liquid markets with rational investors there will be no arbitrage opportunities. Thus what the plots in Figure 5.8 depict is that the expected value of the index at $t = T$ is in general approximated adequately by formula 5.20. The portfolio Value-at-risk and expected shortfall at confidence level α are defined by

$$\mathrm{VaR}_{\alpha,t+\Delta t} = inf\{l \in \mathbb{R} : P(L \leq l) \geq \alpha\} = \overline{F}(\alpha) \tag{5.21}$$

$$\mathrm{ES}_{\alpha,t+\Delta t} = E(L|L \geq VaR_{\alpha,t+\Delta t}) \tag{5.22}$$

where L denotes the random loss, it is assumed that $E(L) < \infty$ and Δt denotes a fixed time horizon. Both of these risk measures have been discussed widely in the literature with the key differences being that ES fulfills the properties of a *coherent* risk-measure, whereas the VaR does not satisfy this property as it lacks subadditivity.[10] Nevertheless VaR is the *de facto* standard for measuring downside risk in finance, e.g. within Solvency II, the regulatory framework relevant to insurance

[10] see Artzner et al., 1999, for a definition of *coherence* in this context.

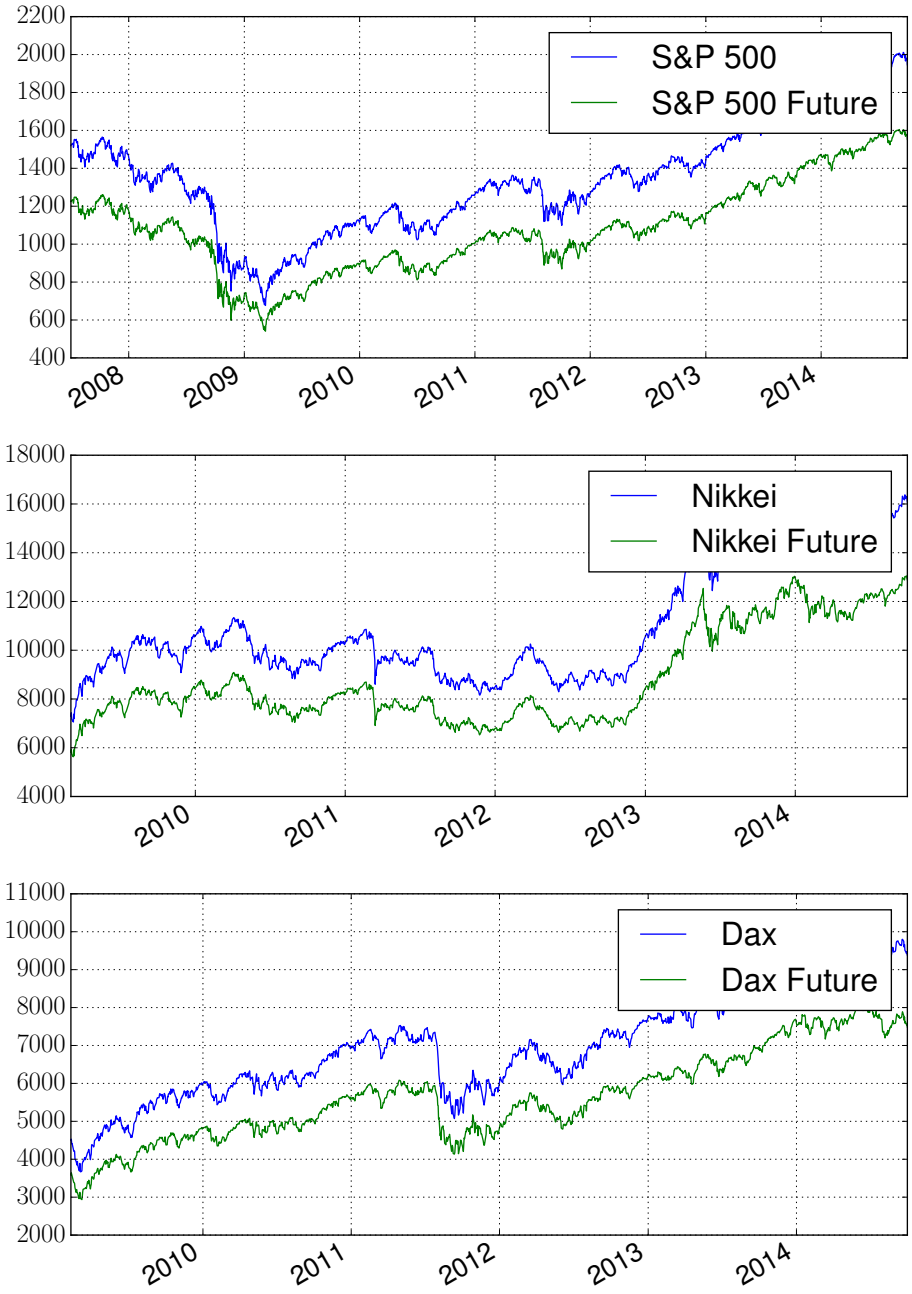

Figure 5.8.: Plots of future-series vs index-series. The future-series was scaled by 0.8 to emphasize the similarities.

companies starting from January 1, 2016 all risks have to be measured by their

$\text{VaR}_{\alpha=99,5}$. We consider the value V in $t = t + 1$ of the following portfolio in Euro:

$$V_{t+1} = \frac{1}{3} P_D e^{R_{1,t+1}} + \frac{1}{3} P_{Ni} e^{R_{2,t+1}} \xi_{\frac{¥}{€}} + \frac{1}{3} P_S e^{R_{3,t+1}} \xi_{\frac{\$}{€}}, \tag{5.23}$$

where $P_D.P_{Ni}$ and P_S denotes the prices of the respective future contract at time t in the base currency, ξ_i, $i \in \{\frac{¥}{€}, \frac{\$}{€}\}$ denote the relevant exchange rates to convert to Euro in price notation and R_i, $i \in \{1, 2, 3\}$ denote the daily returns on the respective futures. The return on the portfolio is consequently given by

$$R^p_{t+1} = \ln\left(\frac{V^p_{t+1}}{V^p_t}\right). \tag{5.24}$$

Figure 5.9 depicts a histogram of the historical returns of this portfolio. The prices

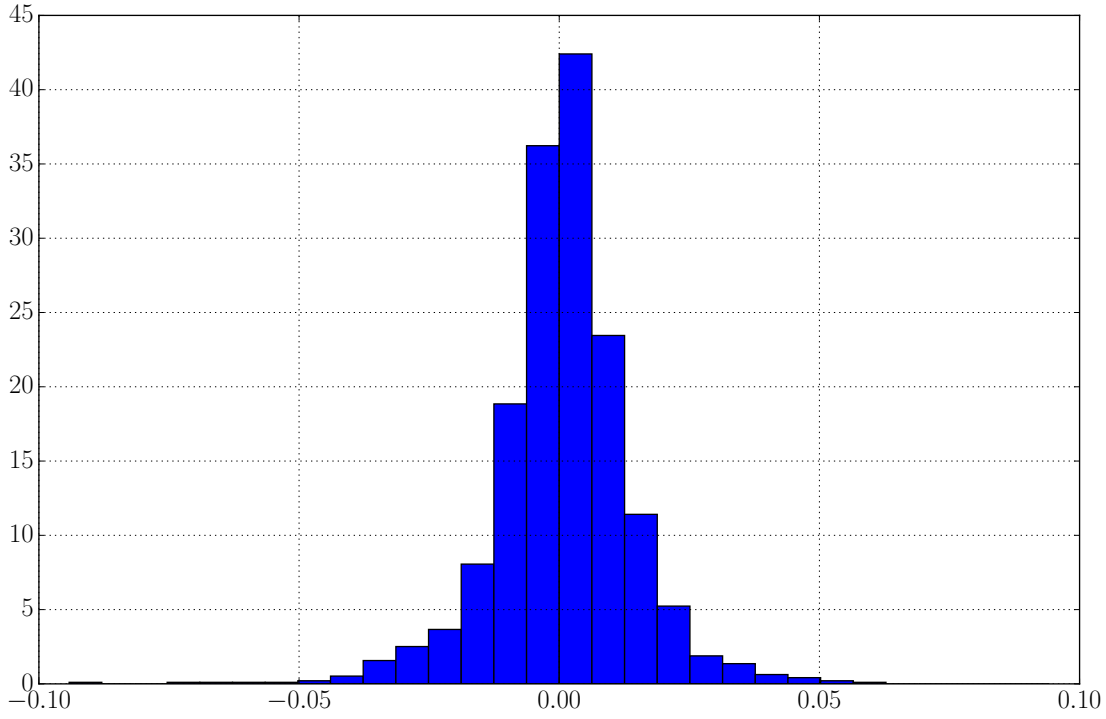

Figure 5.9.: Histogram of portfolio returns.

for the three futures on September 24, 2014 were $P_D = 9590.5€$, $P_{Nik} = 16100¥$ and $P_S = 1991\$$, furthermore we choose to fix the exchange rates at their values from April 30, 2015 for the remainder of this section, i.e. we set $\xi_{\frac{¥}{€}} = 0.0075$ and $\xi_{\frac{\$}{€}} = 0.8919$. We can now proceed to draw samples (r^*_1, r^*_2, r^*_3) from the estimated

future copula and use these samples to estimate the empirical VaR_α and ES_α, where the former is given by the α-quantile of the sample we have generated:

$$\text{VaR}_{t+1} = -r^{*p}_{t+1}, (\alpha(N+1)), \tag{5.25}$$

and the latter by

$$\text{ES}_{t+1} = -\frac{1}{N_{t+1}} \sum_{i=1}^{N} r^{*p}_{t+1,i} \mathbb{I}_{\{x \in \mathbb{R}: x \leq -VaR_{t+1}\}}(r^{*p}_{t+1,i}), \tag{5.26}$$

where N denotes the sample size.

The dependence structure is modeled according to the specifications we introduced in the previous section, i.e. we predict the one week ahead realization of $\rho_\tau^{\max}$at $t = t + 1$, parameterize the LBC accordingly and use this parameterization to generate the necessary samples. These samples have to be transformed to future returns with an appropriate distribution. The obvious similarities between the index series and the future series lead us to consider t-distributions to model the respective future series. Figure 5.10 gives evidence to the adequacy of this specification, while we compare the distributional parameters of the index series and the future series in Table 5.10. The unconditional standard deviation $\hat{\sigma}$ is slightly higher for all index series compared to their future counterparts, as well as the shape parameter $\hat{\nu}$. The latter indicates that the future series seem to possess heavier tails than the index series and if we consider the general shape of the future series we can conclude that they possess more of the leptokurtic characteristics common to financial time series. In contrast, the estimated mean $\hat{\mu}$ is higher for all future series compared to the respective index series.

We further assume that the dependence structure of R_{t+1} does not change within the week.

Table 5.10.: Comparison of distributional parameters Index vs Future

	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\nu}$
Dax	0.000742	0.01929	6.88
Dax future	0.000988	0.00915	3.19
Nikkei	0.000551	0.01653	9.9
Nikkei future	0.000813	0.01081	5.27
S&P 500	0.000901	0.04319	4.71
S&P 500 future	0.000943	0.00721	2.19

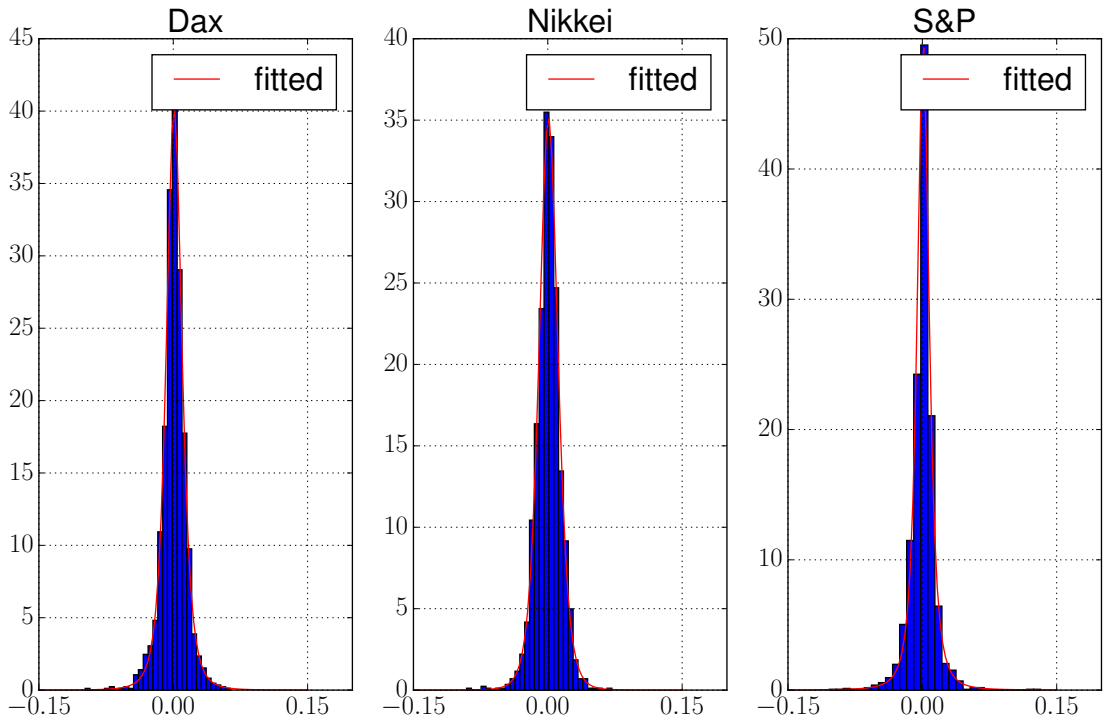

Figure 5.10.: Histograms of fitted t-distributions vs histograms of the future data.

In Table 5.11 we list the last observation of $\rho_\tau^{\max}$at September 24, 2014 and one week ahead predictions up to October 29, 2014. The second row of the table lists the corresponding values of the estimated optimal grid size $\hat{m}_{\mathsf{p}}$.

Table 5.11.: One week ahead predictions of $\rho_\tau^{\max}$starting from t = September 24, 2014 up to October 29, 2014

	$t=0$	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$
$\rho_\tau^{\max}$	0.5999	0.6404	0.6376	0.6351	0.6330	0.6311
$\hat{m}_{\mathsf{p}}$	14	15	15	15	15	15

We see that $\hat{m}_{\mathsf{p}}$ is estimated at 15 for all future dates in consideration due to the solely minor changes in the predicted values of $\rho_\tau^{\max}$. Table 5.12 summarizes the VaR and ES estimates for t = September 25, 2014, for different confidence levels along with the respective confidence intervals. The confidence intervals were estimated via bias-corrected and accelerated bootstrapping, where we have set the number of bootstrap samples B and the individual sample sizes n of each sample $x_b = (x_1, \ldots, x_n), b \in \{1, \ldots, n\}$ to 10000 each.

Table 5.12.: VaR and ES estimates at different confidence levels

α	VaR	95% CI	ES	95% CI
95%	-0.0166	(-0.0174, -0.0159)	-0,0275	(-0.029, -0.0264)
99%	-0.0334	(-0.0352, -0.0323)	-0.0432	(-0.0473, -0.0406)
99.5%	-0.0393	(-0.0422, -0.0374)	-0.0501	(-0.0568, -0.046)

Backtesting Value-at-Risk and expected shortfall estimates

We can now proceed to assess the overall precision of our VaR-estimates and perform a backtesting exercise. We split our dataset into two parts a training and a test set, where the training set is used to train our model, while the test set will be used to perform one step out of sample predictions of VaR_{t+1}. After each estimation we move the training set one step ahead, re-estimate our Index-series GARCH-processes and the one step ahead estimate of $\rho_\tau^{\max}$, which provides us with the necessary parameter to estimate the future Bernstein copula. Starting at March 7, 2014, we proceed in this manner until we arrive at the last date of our test-set, which in our case is April 30, 2015. This gives rise to 275 one step ahead predictions, at each step of which we draw 10000 random samples to estimate VaR_{t+1}.

A note to the computational cost has to be made at this point. Compared to evaluating a closed form functional the computational cost of drawing a large amount of random samples of possibly high dimensional random variables is in general substantially larger for the Bernstein copula. These costs are mainly driven by the optimal grid size $\hat{m}_{\mathsf{p}}$, as the number of grids that need to be evaluated to calculate its density increases in the number of dimensions *as well as* the number of grids i.e. the cost is $\mathcal{O}(\hat{m}_{\mathsf{p}}^d)$, for dimension $d \in \mathbb{N}^+$. Our calculations were mainly conducted on multicore servers with up to 48 processors whose computing power was fully exploited via parallel computing in low-level languages such as C and Fortran. The high number of random samples was chosen to achieve a maximum of precision to avoid the well known curse of dimensionality that non-parametric estimation procedures suffer from. These considerations have to be taken into account for practical purposes in which the practitioner should conduct a cost-benefit analysis before she chooses tho utilize this copula-model. The benefit she gains in utilizing this model lie clearly in the flexibility to estimate an unknown copula arbitrarily precise, as the sample size increases. Figure 5.11 depicts the returns of the portfolio during the specified testing period along with the one-step-ahead forecasts for the VaR_α and the ES_α, for $\alpha \in \{0.9, 0.95, 0.98\}$. The VaR boarder is violated 24, 11 and 1 time respectively, whereas the ES boarder is only violated at the 90% significance level. The estimated Bernstein copula density thus slightly underestimates the VaR as the number of exceedences are slightly less than we would expect under the null

hypothesis. We have placed a dotted line in the same Figure to indicate the mean of the losses that are larger than the VaR_α within the whole sample. The deviation of the mean of the ES-estimates to the observed sample mean is positive at the 95% and the 90% significance level, whereas at the 98% level the deviation is strongest as the sample mean is even lower than the estimated VaR. The VaR and ES estimates thus suggest that the estimated Bernstein copula density assigns slightly too much mass to the tail of the joint distribution of the 3 futures.

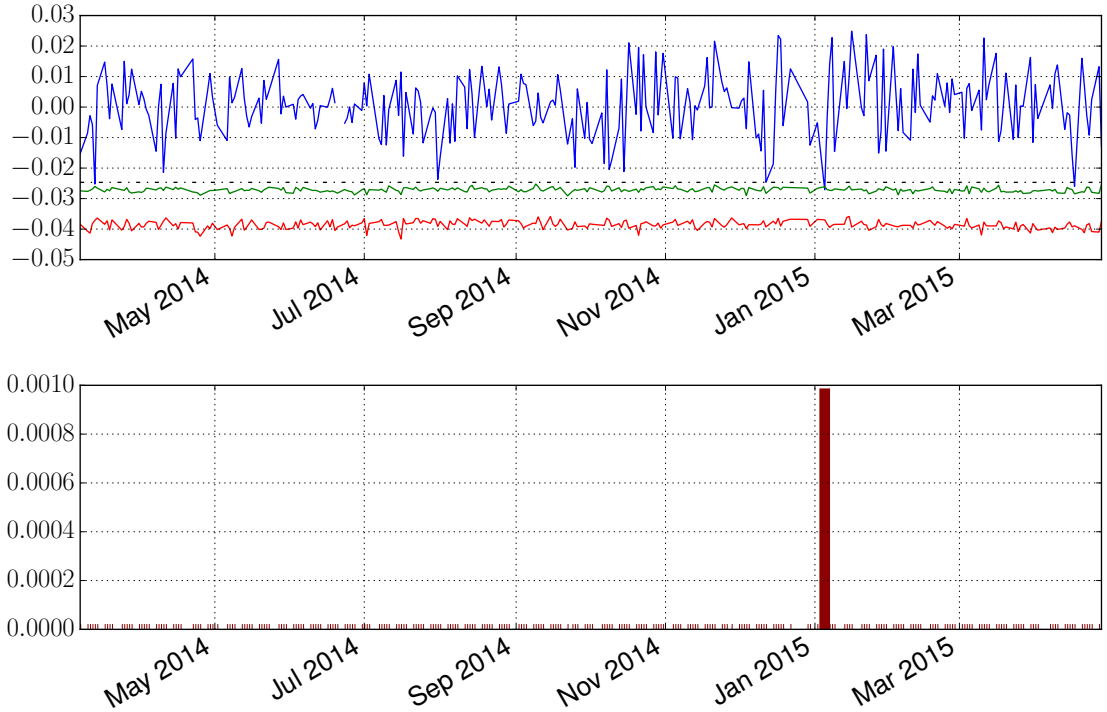

Figure 5.11.: Top: VaR_α (green line) and ES_α (red line) estimates generated from one-step-ahead predictions and realized returns (blue line), $\alpha = 0.98$. The dotted line denotes the empirical mean within the test set of exceedences which are greater than the α-quantile.
Bottom: Plot of the absolute exceedances of VaR_α within the test set, $\alpha = 0.98$.

To add more rigor to these observations we conduct the test of conditional coverage as introduced by Christoffersen, 1998 and Christoffersen and Pelletier, 2004, which is based on the hit sequence

$$h_{t,\alpha} := \begin{cases} 1, & \text{if } R_{p,t} < -VaR_\alpha(R_{p,t}) \\ 0, & \text{otherwise.} \end{cases} \tag{5.27}$$

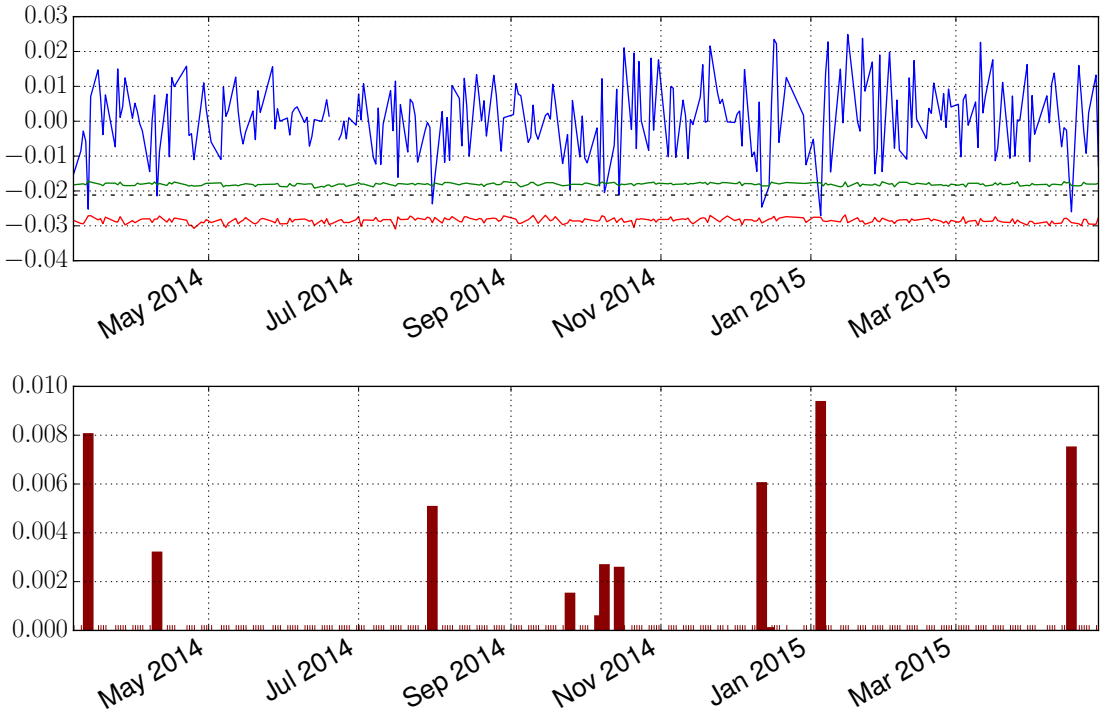

Figure 5.12.: Top: VaR_α (green line) and ES_α (red line) estimates generated from one-step-ahead predictions and realized returns (blue line), $\alpha = 0.95$. The dotted line denotes the empirical mean within the test set of exceedences which are greater than the α-quantile.
Bottom: Plot of the absolute exceedances of VaR_α within the test set, $\alpha = 0.95$.

Under the null hypothesis this sequence follows a Bernoulli distribution with parameter p

$$h_t \sim \text{Bernoulli}(p). \tag{5.28}$$

The alternative hypothesis, based on the observed empirical parameter π is consequently given by

$$h_t \sim \text{Bernoulli}(\pi). \tag{5.29}$$

Thus the test of unconditional coverage tests

$$H_{0,uc} : \pi = p$$

against

$$H_{1,uc} : \pi \neq p,$$

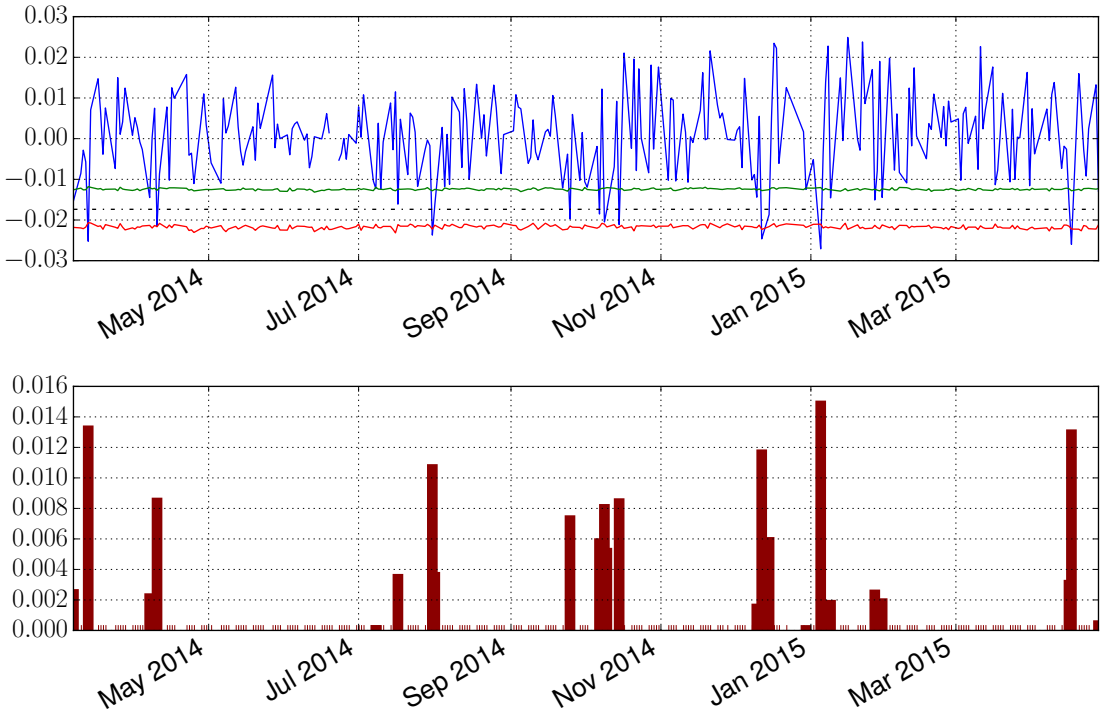

Figure 5.13.: Top: VaR_α (green line) and ES_α (red line) estimates generated from one-step-ahead predictions and realized returns (blue line), $\alpha = 0.9$. The dotted line denotes the empirical mean within the test set of exceedences which are greater than the α-quantile.
Bottom: Plot of the absolute exceedances of VaR_α within the test set, $\alpha = 0.9$.

which is a quite intuitive test for the frequency of the observed VaR-violations. The test for unconditional coverage implicitly assumes that the individual hits are independent. This assumption is tested by the test of independence which assumes that the hit sequence follows a Markov sequence with transition matrix

$$\prod = \begin{pmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{pmatrix}. \tag{5.30}$$

The test is then conducted by testing

$$H_{0,ind} : \pi_{01} = \pi_{11}. \tag{5.31}$$

These two tests can then be combined in a test of conditional coverage, i.e. by

testing the null hypothesis

$$H_{0,cc} : \pi_{01} = \pi_{11} = p. \tag{5.32}$$

For a sequence of T trials the log-likelihood for a Bernoulli variable with unknown parameter π_1 is

$$\ln L(h, p) = \pi_1^{T_1}(1 - \pi_1)^{T-T_1}, \tag{5.33}$$

where T_1, π_1 denote the number of hits and the probability of hits respectively. The maximum likelihood estimate of π is

$$\hat{\pi}_1 = \frac{T_1}{T},$$

which is then used to specify the likelihood ratio test

$$LR_{uc} = -2\left[\ln L(h, \hat{\pi}_1)) - \ln L(h, p_1)\right].$$

The log likelihood for the independence test is

$$\ln L(h, \pi_{01}, \pi_{11}) = (1 - \pi_{01})^{T_- T_{01}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_- T_{11}} \pi_{11}^{T_{11}},$$

with the maximum likelihood estimators for $\hat{\pi}_{01}$ and $\hat{\pi}_{11}$ given by

$$\hat{\pi}_{01} = \frac{T_{01}}{T_0}, \hat{\pi}_{11} = \frac{T_{11}}{T_1}.$$

The test statistics for the test of independence and the test of conditional coverage are then

$$LR_{ind} = -2\left[\ln L(h, \hat{\pi}_{01}, \hat{\pi}_{11}) - \ln L(h, \hat{\pi}_1)\right],$$
$$LR_{cc} = -2\left[\ln L(h, \hat{\pi}_{01}, \hat{\pi}_{11}) - \ln L(h, p_1)\right],$$

respectively. The asymptotic distribution of LR_{uc}, LR_{ind} and LR_{cc}, are chi squared with one degree of freedom for the former ratios and 2 degrees of freedom for the latter.[11] In Table 5.13 we report these test statistics for the current test set along with the theoretical p-values. We can observe that we can't reject the null hypothesis for neither of the three test statistics at the respective significance levels.

In this chapter we introduced the concept of Locally Optimal Bernstein Copulas, which enabled us to adjust our copula density estimatior to specific regions of the sample, which in turn led us to be able to investigate time varying dependence structures from a non-parametric perspective. This concept is sufficiently general

[11] The assumptions underlying the applicability of a chi squared distributions for likelihood are problematic in small sample sizes. The viable alternative is to compute p-values from monte carlo simulations, which we did not conduct, due to the computational cost.

Table 5.13.: Likelihood ratio tests of exceedances at 5 % significance level

α	Exceedances (expected)	Exceedances (realized)	LR_{uc}	LR_{ind}	LR_{cc}
0.02	5.5	1	0.00268 (1.000)	0.00000 (1.000)	0.00269 (1.000)
0.05	13.75	11	0.00000 (1.000)	0.00000 (1.000)	0.00000 (1.000)
0.1	27.5	24	0.00000 (1.000)	0.00000 (1.000)	0.00000 (1.000)

to incorporate different measures of dependence, as long as we make sure that the information captured by the alternative dependence measure is mapped appropriately into the interval $[-1, 1]$. Since $m = 1, 2, \ldots$ only rather large changes in ρ_ξ lead to corresponding changes in $\hat{m}_{\mathfrak{p}}$, i.e. if the local dependence measure differs only slightly from the global measure in the sample the information transported by the locally optimal grid size will not differ from the globally optimal grid size. Nevertheless, if there *are* significant differences in the dependence measures, e.g. in a timeseries context, in times of financial distress, the LBC will adjust accordingly.

6. Summary and Conclusion

The goal of this thesis was to analyze the properties of the Bernstein estimator for copulas and copula densities for real world applications and to improve this estimator methodologically. After introducing the Bernstein copula in Chapter 3 and focusing on its relationship to approximation theory we summarize its large sample properties, which were developed by Sancetta and Satchell, 2004, Bouezmarni et al., 2010, Bouezmarni et al., 2013 and Janssen et al., 2012. We compared the precision of this estimator in a model-selection exercise with medium sized samples, where we showed that in the bivariate case the Bernstein copula density estimator has performance comparable to parametric density estimators. This confirmed the properties we expected theoretically, namely that even though they adjust slowly Bernstein polynomials are fully capable of capturing any dependence structure, while providing the benefit not to be forced to make an assumption with regards to the true underlying process. The latter is indeed crucial since a misspecified copula may lead to a serious bias in the estimation.[1]

While offering enormous flexibility with regard to the functional form it may estimate, the goodness of fit of the estimation is strongly dependent on the specification of the bandwidth parameter m. To address the latter problem we develop a data driven estimator for the bandwidth parameter in section 4.3 of Chapter 4 and analyze its large sample properties. We compare a Bernstein copula parameterized with $\hat{m}$ as derived by this newly developed estimator to a Bernstein copula density parameterized by $\hat{m}$ as derived by the leave one out cross validation estimator introduced in Section 4.1 of Chapter 4. By means of large scale simulation studies with bivariate and trivariate data we find that a Bernstein copula parameterized with our new estimator clearly outperforms the same estimator parameterized with $\hat{m}$ gained from the leave one out cross validation estimator. The application of this estimator also reduces the curse of dimensionality problem as it allows to parameterize a n-dimensional empirical Bernstein copula estimator with its MSE-optimal grid size.

The idea of using information which is specific to a given sample to enhance the performance of the Bernstein copula density estimator is extended to a framework that allows us to adresss several interesting problems known to multivariate density estimation. In this regard we introduce the Locally optimal Bernstein copula LBC

[1] see e.g. Fermanian and Scaillet, 2004.

in Chapter 4, which gives us the tools to adjust our estimation to certain regions of interest, e.g. the tails of a multivariate density. The LBC framework allows us also to capture time varying dependence structures, which to our knowledge has not been applied within a non-parametric setting yet. We assess the applicability of this new procedure in an empirical application for timeseries data, in which we use forecasts of the parameter $\hat{m}$ to obtain an estimate of the future dependence structure within a portfolio constructed from three futures. We calculated VaR and ES-estimates based on the estimated future copula and analyzed the precision of these quantile forecasts via various formal backtests. The results of these backtests indicate that the quality of the forecasts is appropriate since non of the backtests reject the quantiles we obtained.

Within the field of copula estimation many procedures have been analyzed so far, with a clear focus on the estimation of parametric copulas. Only recently there has been an increasing interest in copula estimation via alternative methods such as wavelets, while estimation based on kernel methods or Bernstein polynomials has been known significantly longer. Nevertheless, non-parametric copula estimation takes up only a small part of the literature published in the field, which we can only attribute to the problematic properties in very high dimensions. Being easy to implement the Bernstein copula density estimator provides a viable alternative to the prominent parametric estimators, with appealing properties with regards to its variance and unbiasedness in large samples. But even in small to medium sized samples we find the Bernstein copula estimator to perform reasonably well.

The introduction of a sound estimator for the optimal bandwith parameter $\hat{m}_{\mathsf{p}}$ extends the applicability of the Bernstein copula estimator to estimation problems that have yet been adressed exclusively by parametric copula estimators, such as time varying copula estimation. The concept of exploiting local information to enhance an estimator has already been employed in fields such as regression analysis and density estimation, but only recently have their been applications with a specific focus on multivariate dependence structures.[2]

In this thesis we have focused on analyzing the relationship between the optimal grid size and the strength of dependence as captured by rank correlations. Natural extensions would be to investigate wether estimations in the tails of a multivariate distribution could be enhanced by information as captured by the coefficient of tail dependence or other suitable dependence measures. When utilizing the LBC we have to specify suitable information regions $\Omega_i, i = 1, \ldots, n$, which automatically leads to the question of how to adequately specify these regions. In this regard we would expect pattern recognition tools from machine learning such as the k-Nearest Neighbors algorithm to provide interesting subsets, which could serve as basis for our local estimation procedure.

[2] see Tjøstheim and Hufthammer, 2013.

With regard to the applicability of the empirical Bernstein copula estimator for high dimensional problems, we would suggest to safely utilize this estimator in situations where data scarcity is not an issue. The computational cost may be high but the shape preserving properties of this estimator, especially with respect to the estimation of complicated dependence structures, that cannot easily be reconstructed by parametric copulas, has led us to use to regularly use the Bernstein copula estimator as sound “first guess”, when dealing with copula estimation problems.

A. Appendix

A.1. Additional distance measures: Model-selection

Table A.1.: Mean of $RD_{100,2}$-values from 100 simulations with sample size 100 each. Parameters for the true copula densities:
$(\hat{c}_\theta^{Cl}, \theta = 1)$,$(\hat{c}_\theta^{Fr}, \theta = 5)$,$(\hat{c}_\theta^{t}, \rho = -0.3)$, $(\hat{c}_\theta^{t}, (\rho = 0.65, \nu = 4)$

		True			
	RD_2	$\hat{c}_\theta^{Cl}$	$\hat{c}_\theta^{Fr}$	$\hat{c}_\theta^{Ga}$	$\hat{c}_\theta^{t}$
Estimate	$\hat{c}_B$	0.73397	0.73117	0.75428	0.84533
	$\hat{c}_\theta^{Cl}$	0.63683	0.7206	0.62948	0.87357
	$\hat{c}_\theta^{Fr}$	0.69373	0.72654	0.64523	0.77594
	$\hat{c}_\theta^{t}$	0.64108	0.72683	0.59546	0.65409
	$\hat{c}_\theta^{Ga}$	0.67802	0.72726	0.60862	0.80906

Table A.2.: Mean of $RD_{100,\infty}$-values from 100 simulations with sample size 100 each. Parameters for the true copula densities:
$(\hat{c}_\theta^{Cl}, \theta = 1)$,$(\hat{c}_\theta^{Fr}, \theta = 5)$,$(\hat{c}_\theta^{t}, \rho = -0.3)$, $(\hat{c}_\theta^{t}, (\rho = 0.65, \nu = 4)$

		True			
	$RD_{100,\infty}$	$\hat{c}_\theta^{Cl}$	$\hat{c}_\theta^{Fr}$	$\hat{c}_\theta^{Ga}$	$\hat{c}_\theta^{t}$
Estimate	$\hat{c}_B$	0.99656	0.99791	0.93175	0.99298
	$\hat{c}_\theta^{Cl}$	0.23233	0.35041	0.22501	1.42624
	$\hat{c}_\theta^{Fr}$	0.76547	0.76443	0.53723	0.61474
	$\hat{c}_\theta^{t}$	0.25230	0.54564	0.24187	0.16203
	$\hat{c}_\theta^{Ga}$	0.38189	0.62022	0.25931	0.43027

A.2. Additional descriptive statistics: $\hat{m}_{\mathbf{p}}$ simulation data

Table A.3.: Mean $\overline{x}$ and sample standard deviation s for $x = m_{opt}^{lscv}$ - 2-Dimensional data simulation data. Note that these statistics where calculated with an upper bound of 49 for x that was attained several times, i.e they underestimate their actual values.

	C_{ρ}^{Ga}		$C_{\nu,\rho}^{t}$		C_{θ}^{Fr}	
ρ_τ	$\overline{x}$	s	$\overline{x}$	s	$\overline{x}$	s
-0.9	44.33	4.38	44.33	4.38	21.02	17.06
-0.7	43.96	4.54	44.20	4.39	28.15	17.50
-0.5	41.76	6.55	43.96	4.53	29.13	18.04
-0.3	38.98	10.16	42.96	5.18	34.17	14.48
-0.1	26.22	17.91	42.13	6.53	44.78	4.41
0.1	28.30	16.57	42.00	7.84	25.96	17.32
0.3	36.17	11.31	42.87	6.12	35.37	11.70
0.5	41.78	6.18	44.09	4.41	18.24	15.80
0.7	43.78	5.32	44.33	4.38	43.52	4.55
0.9	44.33	4.38	44.33	4.38	20.17	17.10
ρ_τ	C_{θ}^{Cl}		C_{θ}^{Gu}		C_{θ}^{Jo}	
	$\overline{x}$	s	$\overline{x}$	s	$\overline{x}$	s
0.1	29.17	14.75	21.78	14.99	30.80	14.07
0.3	38.54	9.97	34.76	11.66	41.00	6.82
0.5	43.26	5.53	42.74	6.65	43.83	4.57
0.7	44.17	4.51	44.04	4.52	44.20	4.39
0.9	44.33	4.38	44.33	4.38	44.33	4.38
	C_{ρ}^{Ga}	$C_{\nu,\rho}^{t}$	C_{θ}^{Fr}	C_{θ}^{Cl}	C_{θ}^{Gu}	C_{θ}^{Jo}
CR	54.52	84.13	70.74	29.17	21.78	30.80
MR	158.70	173.87	116.91	81.80	77.50	84.83
OR	176.39	177.17	112.87	88.50	88.37	88.52

Table A.4.: Mean Absolute Scaled Error ($MASE$) across the sample range $[50, 60, \ldots, 500]$ per parameter for 2-dimensional simulated data. The random samples were generated with the parameters that correspond to the respective Kendall's tau values. For C_θ^{Cl} and C_θ^{Gu} negative values of ρ_{tau} either lead to non-strict copulas or, as in the latter case, are not defined.*

MASE							
			$\hat{m}_{\mathsf{p}}$				
ρ_τ	m_{ga}^c	m_t^c	m_{fr}^c	m_{cl}^c	m_{gu}^c	m_{jo}^c	$\frac{1}{n}\sum$
-0.9	0.98	0.92	1.06	n.a.	n.a.	n.a.	0.99
-0.7	1.09	1.49	0.83	n.a.	n.a.	n.a.	1.14
-0.5	5.77	1.44	1.94	n.a.	n.a.	n.a.	3.05
-0.3	2.81	0.30	2.17	n.a.	n.a.	n.a.	1.76
-0.1	14.27	0.04	inf	n.a.	n.a.	n.a.	7.15
0.1	inf	0.09	inf	5.56	8.14	2.02	3.95
0.3	1.96	0.52	102.59	3.71	2.63	3.01	19.07
0.5	4.60	1.15	inf	1.16	5.54	1.73	2.84
0.7	1.39	0.74	1.26	0.60	1.16	0.81	0.99
0.9	1.02	0.70	625.61	0.69	0.48	0.97	104.91
$\frac{1}{n}\sum$	3.77	0.74	105.07	2.34	3.59	1.71	2.09
ρ_τ			$\hat{m}_{opt}^{lscv}$				
ρ_τ	m_{ga}^{lscv}	m_t^{lscv}	m_{fr}^{lscv}	m_{cl}^{lscv}	m_{gu}^{lscv}	m_{jo}^{lscv}	$\frac{1}{n}\sum$
-0.9	8.77	8.77	0.92	n.a.	n.a.	n.a.	6.15
-0.7	8.22	9.27	0.98	n.a.	n.a.	n.a.	6.16
-0.5	5.20	9.22	1.10	n.a.	n.a.	n.a.	5.17
-0.3	4.01	9.88	2.79	n.a.	n.a.	n.a.	5.56
-0.1	1.37	7.01	12.58	n.a.	n.a.	n.a.	6.99
0.1	1.76	8.10	1.73	1.60	0.87	2.48	2.75
0.3	2.84	9.18	3.18	5.40	2.62	7.62	5.14
0.5	6.01	9.77	0.39	7.64	6.22	9.02	6.51
0.7	8.04	9.82	7.87	9.20	8.72	9.27	8.82
0.9	8.77	8.77	0.59	8.77	8.77	8.77	7.40
$\frac{1}{n}\sum$	5.50	8.98	3.21	6.52	5.44	7.43	6.07

* see e.g. Nelsen (2007), an Archimedian copula is considered non-strict when its generator $\phi(0) \neq \infty$. To avoid undefined areas in a non strict Archimedian copulas one has to resign to the notion of a pseudo inverse, when constructing these copulas.

A.3. Additional parameter estimates: Forecasting financial data

Table A.5.: GARCH-M(1,1)$_1$: Ljung-Box Test for autocorrelation in the standardized squared residuals

		Dax	Nikkei	S&P 500
Lag 10	χ^2	4.334	11.239	15.131
	p-value	0.931	0.34	0.127
Lag 15	χ^2	5.895	14.662	15.737
	p-value	0.981	0.476	0.4
Lag 20	χ^2	9.347	17.815	20.278
	p-value	0.979	0.6	0.441

Table A.6.: GARCH(1,1): Ljung-Box Test for autocorrelation in the standardized squared residuals

		Dax	Nikkei	S&P 500
Lag 10	χ^2	4.075	9.905	13.989
	p-value	0.944	0.449	0.173
Lag 15	χ^2	6.377	13.578	14.959
	p-value	0.973	0.558	0.454
Lag 20	χ^2	9.838	16.445	18.824
	p-value	0.971	0.689	0.533

Table A.7.: GARCH(1,1): Parameter estimates for Dax, Nikkei and S&P 500. Robust errors

	GARCH(1,1)				GARCH-M(1,1)$_1$			
	Dax							
	Estimate	Std. Error	t-value	p-value	Estimate	Std. Error	t-value	p-value
μ	0.0007	0.0002	3.0366	0.0024	n.a.	n.a.	n.a.	n.a.
ω	0.0000	0.0000	2.5685	0.0102	0.0000	0.0000	2.5522	0.0107
α	0.1050	0.0203	5.1737	0.0000	0.0989	0.0189	5.2325	0.0000
β	0.8883	0.0177	50.3176	0.0000	0.8936	0.0165	54.2439	0.0000
ν	6.8825	1.0931	6.2963	0.0000	7.1994	1.2029	5.9848	0.0000
	Nikkei							
	Estimate	Std. Error	t-value	p-value	Estimate	Std. Error	t-value	p-value
μ	0.0005	0.0003	1.9407	0.0523	n.a.	n.a.	n.a.	n.a.
ω	0.0000	0.0000	2.6376	0.0083	0.0000	0.0000	2.9548	0.0031
α	0.1032	0.0210	4.9233	0.0000	0.1016	0.0168	6.0526	0.0000
β	0.8785	0.0224	39.2757	0.0000	0.8798	0.0191	46.1400	0.0000
ν	10.0420	2.5550	3.9303	0.0001	10.3108	2.2949	4.4929	0.0000
	S&P 500							
	Estimate	Std. Error	t-value	p-value	Estimate	Std. Error	t-value	p-value
μ	0.0009	0.0002	5.6873	0.0000	n.a.	n.a.	n.a.	n.a.
ω	0.0000	0.0000	3.3437	0.0008	0.0000	0.0000	2.9179	0.0035
α	0.1169	0.0197	5.9237	0.0000	0.1097	0.0174	6.2926	0.0000
β	0.8821	0.0167	52.8220	0.0000	0.8866	0.0159	55.7692	0.0000
ν	4.8024	0.6160	7.7959	0.0000	5.3307	0.7582	7.0311	0.0000

Table A.8.: GARCH-M$(1,1)_2$: Parameter estimates for Dax, Nikkei and S&P 500. Robust errors

	Dax			
	Estimate	Std. Error	t-value	p-value
μ	0.0007	0.0002	3.0386	0.0024
ϕ	-0.2723	0.5832	-0.4670	0.6405
ψ	0.2901	0.5799	0.5003	0.6168
ω	0.0000	0.0000	2.6536	0.0080
α	0.1049	0.0171	6.1372	0.0000
β	0.8884	0.0165	53.9324	0.0000
ν	6.8584	1.1585	5.9201	0.0000
	Nikkei			
	Estimate	Std. Error	t-value	p-value
μ	0.0006	0.0003	1.9637	0.0496
ϕ	-0.3701	0.3708	-0.9983	0.3181
ψ	0.3237	0.3771	0.8584	0.3907
ω	0.0000	0.0000	2.9554	0.0031
α	0.1031	0.0170	6.0491	0.0000
β	0.8786	0.0192	45.7594	0.0000
ν	9.9020	2.1446	4.6172	0.0000
	S&P 500			
	Estimate	Std. Error	t-value	p-value
μ	0.0009	0.0001	6.5568	0.0000
ϕ	0.7974	0.0878	9.0800	0.0000
ψ	-0.8551	0.0756	-11.3097	0.0000
ω	0.0000	0.0000	2.9266	0.0034
α	0.1179	0.0187	6.2978	0.0000
β	0.8811	0.0165	53.4349	0.0000
ν	4.7063	0.5968	7.8864	0.0000

References

Armstrong, J Scott and Fred Collopy (1992). "Error measures for generalizing about forecasting methods: Empirical comparisons". In: *International journal of forecasting* 8.1, pp. 69–80.

Artzner, Philippe et al. (1999). "Coherent measures of risk". In: *Mathematical finance* 9.3, pp. 203–228.

Autin, Florent, Erwan Le Pennec, and Karine Tribouley (2010). "Thresholding methods to estimate copula density". In: *Journal of Multivariate Analysis* 101.1, pp. 200–222.

Baker, Rose (2008). "An order-statistics-based method for constructing multivariate distributions with fixed marginals". In: *Journal of Multivariate Analysis* 99.10, pp. 2312–2327.

Berens, H and RA DeVore (1980). "A characterization of Bernstein polynomials". In: *Approximation Theory* 3, pp. 213–219.

Berentsen, Geir Drage and Dag Tjøstheim (2013). "Recognizing and visualizing departures from independence in bivariate data using local Gaussian correlation". In: *Statistics and Computing*, pp. 1–17.

Bernstein, S.N. (1912). "Démonstration du Théorème de Weierstrass fondée sûr le calcul dés Probabilités". In: *Commun. Soc. Math. Khrakow* 12.2, pp. 1–2.

Berry, Kenneth J and Paul W Mielke Jr (2000). "A Monte Carlo investigation of the Fisher Z transformation for normal and nonnormal distributions". In: *Psychological reports* 87.3f, pp. 1101–1114.

Bouezmarni, T, J Rombouts, and T Taamouti (2008). *A nonparametric test for conditional independence using bernstein density copulas*. Tech. rep. Mimeo.

Bouezmarni, Taoufik, El Ghouch, and Abderrahim Taamouti (2013). "Bernstein estimator for unbounded copula densities". In: *Statistics & Risk Modeling* 30.4, pp. 343–360.

Bouezmarni, Taoufik, Jeroen V. K. Rombouts, and Abderrahim Taamouti (2010). "Asymptotic properties of the Bernstein density copula estimator for alpha-mixing data". In: *J. Multivariate Analysis* 101.1, pp. 1–10.

Boyer, Brian H, Michael S Gibson, and Mico Loretan (1997). *Pitfalls in tests for changes in correlations*. Vol. 597. Board of Governors of the Federal Reserve System.

Brown, Bruce M and Song Xi Chen (1999). "Beta-Bernstein Smoothing for Regression Curves with Compact Support". In: *Scandinavian Journal of Statistics* 26.1, pp. 47–59.

Burnham, Kenneth P and David R Anderson (2004). "Multimodel inference understanding AIC and BIC in model selection". In: *Sociological methods & research* 33.2, pp. 261–304.

Chen, Song Xi and Tzee-Ming Huang (2007). "Nonparametric estimation of copula functions for dependence modelling". In: *Canadian Journal of Statistics* 35.2, pp. 265–282.

Cherubini, Umberto, Elisa Luciano, and Walter Vecchiato (2004). *Copula methods in finance*. John Wiley & Sons.

Christoffersen, Peter F (1998). "Evaluating interval forecasts". In: *International economic review*, pp. 841–862.

Christoffersen, Peter and Denis Pelletier (2004). "Backtesting value-at-risk: A duration-based approach". In: *Journal of Financial Econometrics* 2.1, pp. 84–108.

Claeskens, Gerda, Tatyana Krivobokova, and Jean D Opsomer (2009). "Asymptotic properties of penalized spline estimators". In: *Biometrika* 96.3, pp. 529–544.

Cormier, Eric, Christian Genest, and Johanna G Nešlehová (2014). "Using B-splines for nonparametric inference on bivariate extreme-value copulas". In: *Extremes* 17.4, pp. 633–659.

Cottin, Claudia and Dietmar Pfeifer (2014). "From Bernstein polynomials to Bernstein copulas". In: *Journal of Applied Functional Analysis* 9.

Deheuvels, Paul (1979). "La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance". In: *Acad. Roy. Belg. Bull. Cl. Sci.(5)* 65.6, pp. 274–292.

Devroye, L (1986). *Non-uniform random variable generator*.

Dias, Alexandra and Paul Embrechts (2004). "Dynamic copula models for multivariate high-frequency data in finance". In: *Manuscript, ETH Zurich*.

Diers, Dorothea, Martin Eling, and Sebastian D Marek (2012). "Dependence modeling in non-life insurance using the Bernstein copula". In: *Insurance: Mathematics and Economics* 50.3, pp. 430–436.

Dou, Xiaoling et al. (2014). "EM algorithms for estimating the Bernstein copula". In: *Computational Statistics & Data Analysis*.

Dungey*, Mardi et al. (2005). "Empirical modelling of contagion: a review of methodologies". In: *Quantitative Finance* 5.1, pp. 9–24.

Durrleman, Valdo, Ashkan Nikeghbali, and Thierry Roncalli (2000a). "Copulas approximation and new families". In: *Technical Document*.

– (2000b). *Which copula is the right one*.

Efron, Bradley (1987). "Better bootstrap confidence intervals". In: *Journal of the American statistical Association* 82.397, pp. 171–185.

Embrechts, Paul (2009). "Copulas: A personal view". In: *Journal of Risk and Insurance* 76.3, pp. 639–650.

Embrechts, Paul, Alexander McNeil, and Daniel Straumann (2002). "Correlation and dependence in risk management: properties and pitfalls". In: Cambridge University Press, Cambridge, pp. 176–223.

Engle, Robert (2002). "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models". In: *Journal of Business & Economic Statistics* 20.3, pp. 339–350.

Fermanian, J. "D. and O. Scaillet (2003). Nonparametric estimation of copulas for time series)". In: *Journal of Risk* 5.4.

Fermanian, JD and O Scaillet (2004). *Some statistical pitfalls in copula modelling for financial applications. FAME. Paper No. 108.*

Fermanian, Jean-David, Dragan Radulovic, Marten Wegkamp, et al. (2004). "Weak convergence of empirical copula processes". In: *Bernoulli* 10.5, pp. 847–860.

Forbes, Kristin J and Roberto Rigobon (2002). "No contagion, only interdependence: measuring stock market comovements". In: *The journal of Finance* 57.5, pp. 2223–2261.

Gallant, A Ronald (2009). *Nonlinear statistical models*. Vol. 310. John Wiley & Sons.

Genest, Christian, Esterina Masiello, and Karine Tribouley (2009). "Estimating copula densities through wavelets". In: *Insurance: Mathematics and Economics* 44.2, pp. 170–181.

Gijbels, Ième and Jan Mielniczuk (1990). "Estimating the density of a copula function". In: *Communications in Statistics-Theory and Methods* 19.2, pp. 445–464.

Hyndman, Rob J and Anne B Koehler (2006). "Another look at measures of forecast accuracy". In: *International journal of forecasting* 22.4, pp. 679–688.

Janssen, Paul, Jan Swanepoel, and Noël Veraverbeke (2012). "Large sample behavior of the Bernstein copula estimator". In: *Journal of Statistical Planning and Inference* 142.5, pp. 1189–1197.

Jaworski, P et al. (2010). *Copula Theory and its Applications, volume 198 of Lecture Notes in Statistics-Proceedings.*

Joe, Harry (1990). "Multivariate concordance". In: *Journal of multivariate analysis* 35.1, pp. 12–30.

Jondeau, Eric and Michael Rockinger (2006). "The copula-garch model of conditional dependencies: An international stock market application". In: *Journal of international money and finance* 25.5, pp. 827–853.

Kauermann, Göran, Tatyana Krivobokova, and Ludwig Fahrmeir (2009). "Some asymptotic results on generalized penalized spline smoothing". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2, pp. 487–503.

Kauermann, Göran, Christian Schellhase, and David Ruppert (2013). "Flexible Copula Density Estimation with Penalized Hierarchical B-splines". In: *Scandinavian Journal of Statistics* 40.4, pp. 685–705.

Kim, Gunky, Mervyn J Silvapulle, and Paramsothy Silvapulle (2007). "Comparison of semiparametric and parametric methods for estimating copulas". In: *Computational Statistics & Data Analysis* 51.6, pp. 2836–2850.

Kotz, Samuel and Saralees Nadarajah (2004). *Multivariate t-distributions and their applications*. Cambridge University Press.

Kulpa, Tomasz (1999). "On approximation of copulas". In: *International Journal of Mathematics and Mathematical Sciences* 22.2, pp. 259–269.

Lehmann, EL (1999). *Elements of Large-Sample Theory*. Springer, New York.

Lehmann, Erich Leo and George Casella (1998). *Theory of point estimation*. Vol. 31. Springer.

Li, Xin, P Mikusiński, and Michael D Taylor (1998). "Strong approximation of copulas". In: *Journal of Mathematical Analysis and Applications* 225.2, pp. 608–623.

Li, X et al. (1997). "On approximation of copulas". In: *Distributions with given Marginals and Moment Problems*. Springer, pp. 107–116.

Loader, Clive R et al. (1996). "Local likelihood density estimation". In: *The Annals of Statistics* 24.4, pp. 1602–1618.

Lorentz, George G (2012). *Bernstein polynomials*. American Mathematical Soc.

Lyche, Tom and Knut Morken (2008). "Spline methods draft".

McNeil, Alexander J and Rüdiger Frey (2000). "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach". In: *Journal of empirical finance* 7.3, pp. 271–300.

McNeil, Alexander J, Rüdiger Frey, and Paul Embrechts (2010). *Quantitative risk management: concepts, techniques, and tools*. Princeton university press.

Mittnik, Stefan (2011). "Solvency II calibrations: Where curiosity meets spuriosity". In: *Center for Quantitative Risk Analysis, Department of Statistics, University of Munich, Working Paper* 4.

Morettin, Pedro A et al. (2010). "Wavelet-Smoothed Empirical Copula Estimators". In: *Brazilian Review of Finance* 8.3, pp. 263–281.

Nelsen, Roger B (2007). *An introduction to copulas*. Springer.

Omelka, Marek, Irène Gijbels, Noël Veraverbeke, et al. (2009). "Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing". In: *The Annals of Statistics* 37.5B, pp. 3023–3058.

Patton, Andrew J (2006). "Modelling asymmetric exchange rate dependence*". In: *International economic review* 47.2, pp. 527–556.

– (2009). "Copula–based models for financial time series". In: *Handbook of financial time series*. Springer, pp. 767–785.

Petrone, Sonia (1999a). "Bayesian density estimation using Bernstein polynomials". In: *Canadian Journal of Statistics* 27.1, pp. 105–126.

– (1999b). "Random bernstein polynomials". In: *Scandinavian Journal of Statistics* 26.3, pp. 373–393.

Petrone, Sonia and Larry Wasserman (2002). "Consistency of Bernstein polynomial posteriors". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64.1, pp. 79–100.

Pfeifer, Dietmar, D Strassburger, and J Philipps (2009). "Modelling and simulation of dependence structures in nonlife insurance with Bernstein copulas". In: *Preprint*.

Qu, Leming, Yi Qian, and Hui Xie (2009). "Copula density estimation by total variation penalized likelihood". In: *Communications in Statistics-Simulation and Computation* 38.9, pp. 1891–1908.

Qu, Leming and Wotao Yin (2012). "Copula density estimation by total variation penalized likelihood with linear equality constraints". In: *Computational Statistics & Data Analysis* 56.2, pp. 384–398.

Sancetta, A. and S.E. Satchell (2001). *Bernstein Approximations to the Copula Function and Portfolio Optimization*. Tech. rep. 0105,

Sancetta, Alessio and Stephen Satchell (2004). "The Bernstein copula and its applications to modeling and approximations of multivariate distributions". In: *Econometric Theory* 20.03, pp. 535–562.

Schmid, Friedrich and Rafael Schmidt (2007). "Multivariate extensions of Spearman's rho and related statistics". In: *Statistics & Probability Letters* 77.4, pp. 407–416.

Scott, David W (2012). "Multivariate density estimation and visualization". In: *Handbook of Computational Statistics*. Springer, pp. 549–569.

Sheather, Simon J et al. (2004). "Density estimation". In: *Statistical Science* 19.4, pp. 588–597.

Shen, Xiaojing, Yunmin Zhu, and Lixin Song (2008). "Linear B-spline copulas with applications to nonparametric estimation of copulas". In: *Computational Statistics & Data Analysis* 52.7, pp. 3806–3819.

Tenbusch, Axel (1994). "Two-dimensional Bernstein polynomial density estimators". In: *Metrika* 41.1, pp. 233–253.

– (1997). "Nonparametric curve estimation with Bernstein estimates". In: *Metrika* 45.1, pp. 1–30.

Tjøstheim, Dag and Karl Ove Hufthammer (2013). "Local Gaussian correlation: A new measure of dependence". In: *Journal of Econometrics* 172.1, pp. 33–48. URL: `http://www.sciencedirect.com/science/article/pii/S0304407612001741`.

Van der Vaart, Aad W (2000). *Asymptotic statistics*. Vol. 3. Cambridge university press.

Vitale, Richard A (1975). "A Bernstein polynomial approach to density function estimation". In: *Statistical inference and related topics* 2, pp. 87–99.

Weiss, Gregor NF and Marcus Scheffer (2012). "Smooth nonparametric bernstein vine copulas". In: *Available at SSRN 2154458*.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. .5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

Rose, Doro

Name, Vorname

Ort, Datum

Unterschrift Doktorand/in

Formular 3.2